ARI Research Note 96-38

Bending the Power Law: The Transition From Algorithm-Based to Memory-Based Performance

Timothy Charles Rickard
University of Colorado

Research and Advanced Concepts Office
Michael Drillings, Acting Director

March 1996

ARMY RESEARCH INSTITUTE

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract for the Department of the Army

University of Colorado

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

BENDING THE POWER LAW: THE TRANSITION FROM ALGORITHM-BASED TO MEMORY-BASED PERFORMANCE

by

TIMOTHY CHARLES RICKARD

B.S., University of Alabama, 1987

M.S., University of Alabama, 1989

M.A., University of Colorado, Boulder, 1992

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology

1994

This thesis for the Doctor of Philosophy degree by

Timothy Charles Rickard

has been approved for the

Department of Psychology

by

Lyle E. Bourne Jr.

Alice F. Healy

Date 13, July 1994

REPORT DOCUMENTATION PAGE

1. REPORT DATE 1996, March	2. REPORT TYPE Annual	3. DATES COVERED (from. . . to) August 1993-August 1994
4. TITLE AND SUBTITLE Bending the Power Law: The Transition From Algorithm-Based to Memory-Based Performance		5a. CONTRACT OR GRANT NUMBER MDA903-93-K-0010 5b. PROGRAM ELEMENT NUMBER 0601102A
6. AUTHOR(S) Timothy Charles Rickard (University of Colorado)		5c. PROJECT NUMBER B74F 5d. TASK NUMBER 2901 5e. WORK UNIT NUMBER C07
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Colorado Campus Box B-19 Boulder, CO 80309		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: PERI-BR 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI 11. MONITOR REPORT NUMBER Research Note 96-38

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

COR: George Lawton

14. ABSTRACT *(Maximum 200 words)*:

Two theories of tasks that exhibit a transition from algorithm-based to retrieval-based performance were compared. The instance theory of automaticity assumes parallel strategy execution and instance-based memory representation and predicts power function reduction in mean reaction time and standard deviation of mean reaction time (SD) with practice. An alternative theory is proposed that assumes nonparallel strategy execution and strength-based memory representation and that predicts, among other things, power function speed-up and reduction in SD within each strategy, and systematic deviations from power functions in both of these variables when strategy transition occurs. Two of these experiments employing pseudoarithmetic are reported. The results of these experiments are consistent with the new theory that assumes nonparallel strategy execution and strength-based memory representation. These results also constitute the first convincing demonstration of a class of adult skill acquisition tasks for which the power law of practice does not apply overall, a finding that should have notable implications for a variety of human skill acquisition theories.

15. SUBJECT TERMS

Mental arithmetic   Automaticity   Power law of practice   Training

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	120	

Rickard, Timothy Charles (Ph.D., Psychology)

Bending the Power Law: The Transition from Algorithm-based to Memory-based Performance

Thesis directed by Professor Lyle E. Bourne, Jr.

Two theories of tasks which exhibit a transition from algorithm-based to retrieval-based performance were compared. The instance theory of automaticity assumes parallel strategy execution and instance-based memory representation and predicts power function reduction in mean reaction time (RT) and standard deviation of mean reaction time (SD) with practice. An alternative theory is proposed which assumes nonparallel strategy execution and strength-based memory representation and which predicts, among other things, power function speed-up and reduction in SD within each strategy, and systematic deviations from power functions in both of these variables when strategy transition occurs.

In two experiments employing pseudo-arithmetic tasks subjects exhibited a complete transition from using an algorithm to retrieving answers to individual problems directly from memory (as revealed by both strategy probing and RT data) on or before the 60th exposure to each problem. Power function speed-up and reduction in SD with practice clearly did not hold for the overall data in either experiment, but did hold generally within each strategy. Transfer data from Experiment 1 also indicated that learning was highly specific to the practiced problems. Overall these results are consistent with the new theory which assumes nonparallel strategy execution and strength-based memory representation. These results also constitute the first convincing demonstration of a class of adult skill acquisition tasks for which the power law of practice does not apply overall, a finding which should have notable implications for a variety of human skill acquisition theories.

DEDICATION

This dissertation is dedicated to my parents, Charlie and Barbara Rickard, whose support and encouragement throughout graduate school was indispensable.

ACKNOWLEDGEMENTS

Many thanks to my committee chair, Lyle Bourne, and to committee members Alice Healy, Mike Mozer, Reid Hastie, and Walter Kintsch. Thanks also to Anders Ericsson, Bob Rehder, and Debbie Clawson and Steve Romero for many interesting discussions. I am also grateful to Steve and to Jodi Schaeffer for their help in collecting data.

CONTENTS

TABLES

Table

FIGURES

Figure

CHAPTER 1

INTRODUCTION

Performance on virtually any task can be improved by practice. Indeed, in some cases, practice can produce an order of magnitude or greater reduction in task completion time. Well known examples include typing, solving mental arithmetic problems (Siegler, 1988), reading inverted text (Kolers, 1975), and solving geometry proofs (Neves & Anderson, 1981). The striking improvement with practice on these tasks suggests fundamental qualitative changes in the psychological processes underlying performance, and identification of the mechanisms which support these changes is one of the most interesting and important challenges for researchers exploring skill acquisition. This study contributes to this line of research by exploring a class of tasks which exhibits a transition from use of a multi-step algorithm to single-step retrieval of answers from memory. A classic example is basic mental arithmetic (e.g., 4 + 7 = ?; 5 x 6 = ?). During initial stages of learning, children often use a variety of laborious counting algorithms which can require 10 to 20 seconds to execute (Siegler, 1988). With practice, however, they learn to retrieve answers to individual problems directly from memory. By adulthood, the direct retrieval strategy typically yields answers on the order of 1 second or faster.

The primary purpose of this study is to test two candidate theories of the transition from algorithm-based to retrieval-based performance. One of these, the instance theory of automaticity (Logan, 1988), incorporates two fundamental assumptions: (a) each exposure to an item establishes an independent episode, or instance, of that event in memory, and (b) the algorithm and retrieval strategies are executed in parallel on each trial. A new theory which I will introduce makes two diametrically opposing assumptions: each exposure to an item strengthens a prototype representation of that item, and one and only one strategy can be executed at any given time. These issues of strength- versus instance-based memory representation and of parallel versus serial information processing are of course not new to psychology, and current evidence suggests that either of the alternatives for both issues may be correct across various domains of information processing. For example, parallel processing is widely believed to support a variety of perceptual and memory processes (e.g., McClelland & Rumelhart, 1981). On the other hand, there is strong evidence that information processing in more complex domains such as problem solving has fundamentally serial components (e.g., Anderson, 1983). With respect to memory representation, it seems clear that episode-like memory processes are operative in some circumstances, such as recall of highly contextualized life events, but there is also reason to believe that strengthening of a prototype representation takes place in some situations. As I will discuss in detail below, tasks exhibiting a transition from algorithm to retrieval appear to provide a promising context in which to explore further these two central issues of human information processing.

A second and more empirically-based motivation for this paper is to explore in detail the functional characteristics of speedup in reaction time (time to produce a correct answer to any problem) with practice on this class of tasks. The power law of practice (Newell & Rosenbloom, 1981) predicts a smooth and negatively accelerating reduction in reaction time. Tasks which exhibit a transition from algorithm to retrieval often exhibit an order of magnitude or greater reduction in reaction time over a relatively short practice interval, and thus should provide a sensitive test of this law.

The questions of whether the power law holds for this task domain and which of the theoretical perspectives outlined above is more appropriate turn out to be closely related. To preview, the instance theory predicts that speedup with practice will follow the power law. In contrast, the new theory proposed in this dissertation predicts that the power law holds within a given strategy (e.g., algorithm or retrieval), but that systematic deviations from the power law will be present when strategy transitions occur. Both of these theories, and two experiments designed to differentiate between them, will be discussed in more detail in later sections. I will begin by overviewing the power law of practice, and by reviewing briefly the literature on tasks exhibiting a transition from algorithm-based to memory-based performance.

The Power Law of Practice

Power function speedup with practice has been observed across a wide variety of tasks, including retrieval of facts from memory (Pirolli & Anderson, 1985; Rickard, Healy, & Bourne, in press), repeating sentences (MacKay, 1982), proving geometry theorems (Neves & Anderson, 1981), learning editing routines (Moran, 1980), rolling cigars (Crossman, 1959), and evaluating logic circuits (Carlson, Sullivan, & Schneider, 1989). In fact, power function speedup appears to be so ubiquitous that Newell and Rosenbloom (1981) conferred on it the status of a scientific law. The power function predicts a negatively accelerating rate of speedup as a function of practice. That is, it predicts substantial speedup from trial-to-trial during early stages practice, but progressively less speedup from trial-to-trial during later stages. In formal terms,

$$RT = a + b*N^{-c},$$

where *RT* is the time required to do the task, *N* is the number of practice trials, and *a*, *b*, and *c* are parameters. The term *b*N**-c* goes to zero as *N* goes to infinity, and thus the parameter *a* represents the asymptotic RT. The parameter *b* is the difference between the RT on the first trial, and the RT at asymptote, and *c* is a rate parameter which determines how quickly the RT approaches asymptote. A simplified version of the power function which ignores the asymptote,

$$RT = b*N^{-c},$$

fits RT data essentially as well in most circumstances (see Newell & Rosenbloom, 1981).

The power function is linear when plotted in log-log coordinates provided that the asymptote is first subtracted. Thus,

$$\log(RT - a) = \log(b) - c*\log(N).$$

This log-log linearity can be a powerful diagnostic tool in evaluating how closely data conform to a power function. Often substantial and systematic deviations from linearity in log-log plots can be detected visually even when statistical regressions fits yields r^2 values of .9 or higher. Thus, in evaluating a power function fits to data, both statistical measures and visual inspections of log-log plots are needed (Newell & Rosenbloom, 1981).

The power law has been the most important empirical constraint influencing the development of a variety of skill theories, including those of Anderson (1983, 1993), Cohen, Dunbar and McClelland (1990), Logan (1988), MacKay (1982), and Newell and Rosenbloom (1981), and it is generally believed to hold for any task domain. Logan (1988), for example, describes the power law as a "benchmark prediction that theories of skill acquisition must make to be serious contenders." Despite such strong conclusions, there is currently little empirical evidence in support of the assumption that the law holds for tasks exhibiting a transition from algorithm to retrieval. Indeed, as I will discuss later, the available data hint at the possibility that the law does not hold for this task domain. One of the purposes of the current research is to collect new data which will more decisively address this question.

The Transition from Algorithm-based to Memory-based performance

A prototypical example of the transition from algorithm-based to retrieval-based performance is children's arithmetic (e.g., 4 + 5 = ?; 3 x 7 = ?). Several investigators

(Siegler, 1988; Siegler & Jenkins, 1989) have shown that children initially use a generic counting algorithm to solve these problems. For example, they may solve 4 + 5 by starting with 5 and counting on four times to reach 9. When learning multiplication, they often use a repeated adding algorithm, such that, for example, 4 x 7 is solved by executing the algorithm 7 + 7 + 7 + 7. With practice there is a transition to retrieval of individual addition and multiplication facts directly from memory. By adulthood, single-step retrieval of answers to both addition and multiplication problems is pervasive (for a review see Ashcraft, 1992).

The transition from algorithm to retrieval has also been observed in adults. In Logan's (1988) alphabet arithmetic task, college subjects are asked to verify equations of the form A + 2 = C, B + 3 = E, and D + 4 = G (see also Compton & Logan, 1991; Klapp, Boches, Trabert, & Logan, 1991; Logan & Klapp, 1991; Logan, 1992). The algorithm for determining whether the provided answer is true or false involves starting with the first (left-hand) letter, counting sequentially through the alphabet the number of times indicated by the digit (or addend), and comparing the letter arrived at to the letter presented to the right of the equal symbol. Thus, the first and second equations above are true, and the third equation is false. During initial practice, subjects typically use the algorithm defined above. Later, there is a transition to direct retrieval of answers from memory (Compton & Logan, 1991; Logan & Klapp, 1991). The empirical evidence confirming this effect involved a comparison of the effects of addend size on RTs at the beginning and at the end of practice (see Compton & Logan, 1991; Logan & Klapp, 1991). For addend sizes ranging from 2 to 5, there was a substantial increase in RT with increasing addend size at the beginning of practice. This effect is to be expected because the counting algorithm requires more steps with increasing addend size. However, there were no systematic RT increases with addend size after several sessions of practice. If practice had simply resulted in faster execution of the algorithm, the addend size effect would not be expected to disappear. The absence of a systematic addend effect is consistent, however, with the claim that subjects learned to solve the problems using single-step access to memory.

A related type of strategy transition, from mediated retrieval to direct or single-step retrieval, has also been observed in adults using the keyword method of second language vocabulary learning (Crutcher, 1989). Retrieval of the foreign language equivalent of a native language word using the keyword method involves two steps: (a) associating a foreign word (e.g., *doronico*, which means leopard in Spanish) with a phonologically similar native language keyword (e.g., *door*), and (b) forming an image linking the keyword with the native equivalent of the foreign word (e.g., imagining a leopard walking through a door). Using retrospective protocols, Crutcher showed that with sufficient practice, subjects often no longer consciously used the mediating keyword to retrieve the English equivalent. That is, subjects began accessing the English word directly upon presentation of the foreign word.

Although the research to date demonstrating a transition from algorithm-based to memory-based performance has mostly been conducted in the laboratory, it is likely that this phenomenon also occurs in natural contexts. Children's mental arithmetic discussed above is an obvious example, and a variety of other tasks occurring in both educational and workplace settings almost certainly exhibit similar learning phenomena. It is also plausible, based on the laboratory work exploring the keyword method (Crutcher, 1989), that any mnemonic which is initially used to recall verbal information is, with sufficient experience, replaced by direct retrieval of the desired information. Thus, the transition from algorithm-based to retrieval-based performance is probably a quite common psychological phenomenon.

Although the research outlined above leaves little doubt that the transition to retrieval occurs across a variety of contexts, many basic questions regarding the psychological processes which mediate this phenomenon have not been conclusively addressed. Is the memory representation which supports direct retrieval best understood as instance-based or as strength-based? How is the development and character of memory in this domain similar to or different from that of other domains? Does the development of associations which support memory retrieval occur as an automatic consequence of practice, or is a strategic effort to development of such associations critical? Are the algorithm and retrieval strategies the only qualitatively distinct strategies? Are all strategies executed in parallel, with the first strategy to produce the answer determining performance, or must strategies be executed serially? If only one strategy can be executed at a time, then how is this strategy selected? To what extent will the answers to these questions depend on parameters of the task or on individual differences? The two candidate theoretical frameworks discussed below make very different predictions regarding most of these questions.

The Instance Theory of Automatization

Logan's (1988, see also Compton & Logan, 1991; Logan, 1992) instance theory of automatization incorporates three basic assumptions. First, it assumes that "encoding into memory is an obligatory, unavoidable consequence of attention" (Logan, 1988, p. 493). Second, it assumes that "retrieval from memory is an obligatory, unavoidable consequence of attention" (Logan, 1988, p. 493). Third, it assumes that "each encounter with a stimulus is encoded, stored, and retrieved separately" (Logan, 1988, p. 493). This last assumption makes the theory an instance theory of memory, which contrasts it with a variety of strength-based theories of memory processes (e.g., Anderson, 1983; MacKay, 1982).

Three additional assumptions allow for derivation of a quantitative model which can be applied directly to data from tasks exhibiting a transition from algorithm to retrieval (see Logan, 1988, for a detailed discussion). First, the algorithm and each memory instance are assumed to compete in parallel, and independently, on each trial. The process which finishes the race first controls the response. Second, each episode, or instance, has the same distribution of finishing times which does not change with practice. Third, the algorithm has a separate distribution of finishing times which does not change with practice. The memory strategy comes to dominate the race as practice proceeds because as more memory episodes accrue the probability that one of them will win the race continually increases.

Using a combination of formal mathematical proofs and Monte Carlo simulations, Logan (1988) showed that the instance theory predicts that (a) the speedup in mean reaction time (RT), as well as the reduction in standard deviation (SD) of the reaction time, follows a power function of practice, and (b) the the rate parameters for the speedup in mean RT and reduction in SD are the same. Expressed as equations, the instance theory's predictions for the RT and SD are:

$$RT = a_1 + b_1 * N^{-c}$$

$$SD = a_2 + b_2 * N^{-c}.$$

Logan (1988) showed that, given special assumptions about the form of the reaction time distributions for the algorithm and for retrieval, the instance theory also predicts that the probability of using the algorithm decreases as a power function of practice. He did not fit power functions to the algorithm probability data that were generated in the Monte Carlo simulations, however, and he did not state that his theory predicts that the probability of using the algorithm deceases as a power function in the general case (i.e., across a wide range of candidate RT distributions). He did, however, provide plots of the transition data that were generated by Monte Carlo simulations (See Logan, 1988, Figure 3), and it is clear from these plots that the probability of using the algorithm is predicted to be a negatively accelerating function of practice which closely resembles the power function.

Logan (1988, Experiment 4) tested the instance theory using the alphabet arithmetic task described previously. Each subject received 72 blocks of practice, across 12 sessions, on 10 true and 10 false problems at each of four levels of addend size (2, 3, 4, and 5), for a total of 80 problems per block. Figure 1, reprinted from Logan (1988) shows the results for true alphabet arithmetic equations at each level of addend size. (The results for false problems were equivalent). The data are plotted in log-log coordinates, and the straight lines are best fits of the instance theory. The fits overall are reasonably good. However, on close examination it is clear that the model underestimates the RTs and SDs during the middle portion of practice, and it overestimates these values toward the end of practice. This trend is weak for addend sizes of 2 and 3, but is clearly apparent for addends sizes of 4 and 5. For equations with addend of 5, for example, the instance theory fits consistently underestimate the RT by more than 300 ms (in terms of anti-log means) around the middle of the practice interval, and overestimates the RT by about the same amount by the end of practice. Logan (1988) acknowledged these deviations, but argued that they do not constitute a serious challenge to the instance theory for two reasons. First, no existing model of skill acquisition predicts the deviations (because all current theories predict power-function speedup) and thus evidence against the instance theory is also evidence against the other models. Second, Logan suggested a post hoc modification to the instance theory which improves the fit to the data. Some subjects reported at the end of the experiment that they used special mnemonics to deal with the problems with addends of 5. Logan proposed that subjects shifted to using mnemonics between the fourth and fifth sessions of practice, and that the use of mnemonics resulted in more efficient, or more memorable, traces, with a faster associated RT distribution. A modified version of the instance theory which incorporates this assumption can account for the bulk of the deviations from the power functions observed in the addend 5 data.

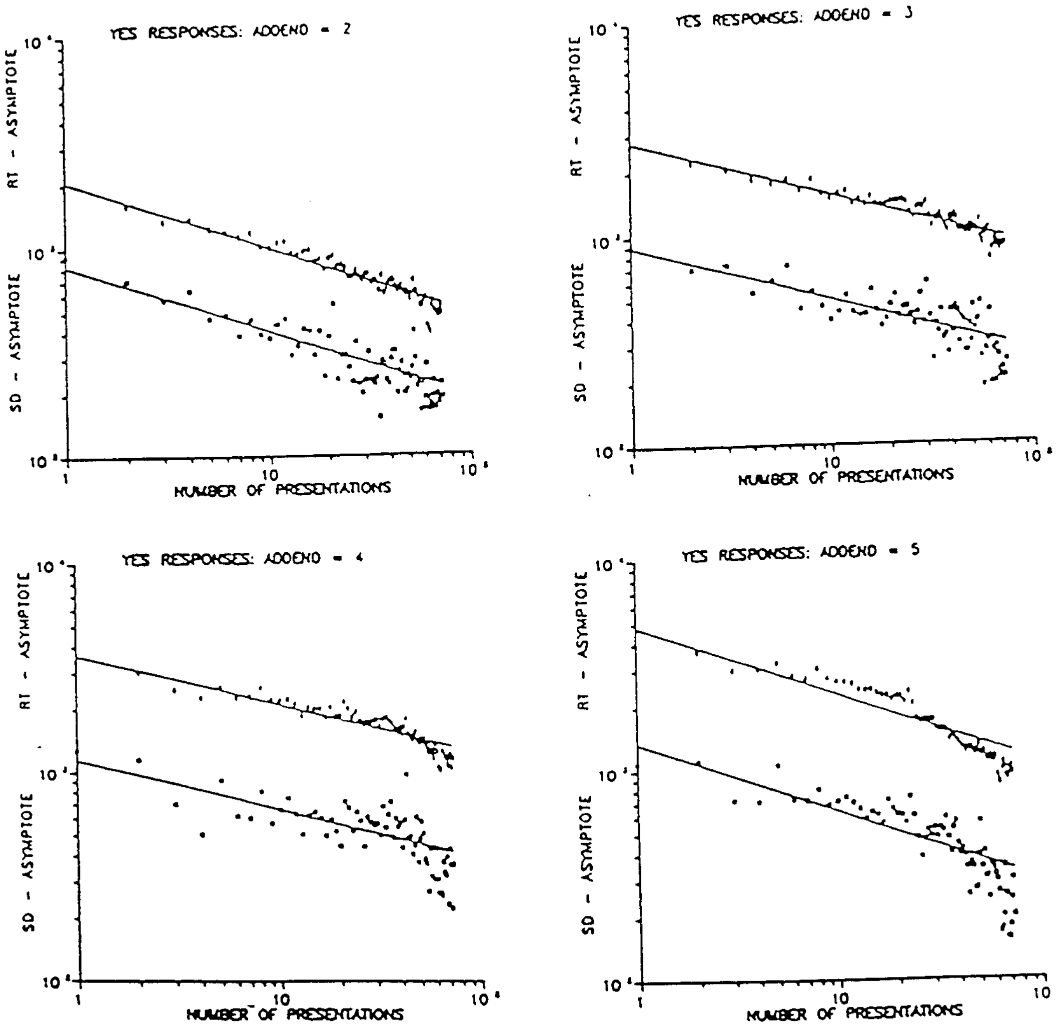


Figure 1. Log means and standard deviations of reaction times to verify true alphabet arithmetic equations as a function of the log of the number of presentations and the magnitude of the addend size. Lines represent fitted power functions constrained to have the same exponent for means and standard deviations. Reprinted from Logan (1988).

A Nonparallel Strategies Theory
of the Transition from Algorithm to Retrieval

A major feature of the instance theory is the assumption that algorithm and retrieval strategies can be executed in parallel and independently of one another. An alternative and intuitively plausible assumption is that strategy execution is always nonparallel, such that one and only one strategy can be executed at any given time. In this section I sketch a preliminary theory, which I will refer to simply as the nonparallel strategies (NPS) theory, which embodies this idea in information processing language. There are two motivations for this development. First, expression of the notion of nonparallel strategy execution in information processing terms will facilitate a subsequent comparison of the empirical and conceptual evidence for this approach in comparison to the instance theory. Second, the theory will provide the foundation for development of a more precise quantitative model to be introduced later.

The NPS theory makes four straightforward assumptions which are more or less consistent with several general theories of human information processing (e.g., Anderson, 1983; Newell, 1993). First, the flow of information processing on tasks exhibiting a transition from algorithm to retrieval is assumed to be strongly goal-directed; performance requires a sequence of changing goals and subgoals which focus attention and which guide execution of the steps of a complex strategy such as an algorithm. A corollary to this assumption is that there is a general goal of "producing the answer" at the outset of every problem presentation, and that, soon after problem presentation, a separate subgoal either to "execute the algorithm" or to "execute retrieval" is selected. If the algorithm subgoal is selected, additional subgoals which guide performance through the algorithm are subsequently selected. If the retrieval subgoal is selected, then direct retrieval of the answer from memory will take place.

The second and closely related assumption is that attention can be focused on at most one goal or subgoal, which I will refer to as the focus goal, at any time. The third assumption is that the focus goal places deterministic constraints on what can be retrieved from memory. Information can be retrieved from long-term memory only if it is associated with information currently in working memory and is consistent with the current focus goal.

Assumptions 1 through 3 effectively constrain the NPS theory to predict one-at-a-time strategy execution. If, for example, the algorithm subgoal is selected, then direct retrieval is precluded because attention can be focused on only one goal at a time and because only information which is consistent with the focus goal can be retrieved from memory. The fourth assumption of the theory is that memory is strength-based. Each learning event for a given item strengthens a generic association which is "abstracted" across repetitions. The probability of using the retrieval strategy is determined solely by the strength of the association between the problem and the answer. Some type of conflict resolution is assumed which allows the retrieval subgoal to be selected once the strength of the association between the problem and the answer reaches some minimal strength value.

Conceptual and Empirical Evidence Bearing on

the Assumptions of the Instance and NPS Theories

Assumptions 1, 2 and 3 of the NPS theory lead to the prediction of nonparallel strategy execution. One and only one strategy can be executed at any given time. At least one of these assumptions must be inconsistent with the instance theory, because that theory assumes parallel strategy execution. Because goal structures are never directly considered in the development of the instance theory (Logan, 1988), the exact nature of the inconsistency of these assumptions with the theory is unclear. At a general conceptual level, however, there appear to be two distinct possibilities. First, consider Assumption 3 of the NPS theory, which states that information associated with an item can be retrieved only when the item is being attended and when the information is consistent with the focus goal. Relaxing this assumptions in the following way might conceivably allow for parallel strategy execution: attention to the general goal (such as "solve the problem") could be assumed to be sufficient to start the retrieval process, which then continues ballistically even after attention is shifted to other focus goals as necessary to execute the algorithm. This assumption appears to be consistent with the instance theory assumption that "retrieval is an unavoidable, obligatory consequence of attention." However, evidence from Zbordoff and Logan (1986) indicates that this sort of ballistic or "obligatory" retrieval does not typically occur. Using a mental arithmetic task, Zbordoff and Logan (1988) showed that retrieval processes are initiated in an obligatory fashion, but that, if tasks goals are altered, the retrieval process is aborted before the answer is generated either overtly or covertly.

Now consider the possible effects of relaxing the second assumption of the NPS theory that attention to a focus goal is all-or-none. An obvious alternative is that attention is divided between the two strategies throughout the task. This assumption is inconsistent with Logan's (1988) claim that the instance theory circumvents the need for assumptions about resource allocation, but it is not in principle inconsistent with any other aspect of the model. The experimental literature on attention, however, indicates that even if attention can be divided (and there is controversy on whether this effect has been demonstrated in contexts analogous to the current one, see Shiffrin, 1988), it is divided neither easily nor naturally. Also, in studies in which divided attention has putatively been demonstrated, subjects are explicitly instructed to attend to two or more tasks. In contrast, in tasks exhibiting a transition from algorithm to retrieval, subjects are not instructed to attend to both strategies (indeed, they are not instructed at all regarding attentional allocation). Thus, it is more reasonable to assume that they typically choose to attend to one strategy at a time. Further, in the event that the retrieval strategy is not selected at the onset of a trial, there is little motivation for the subject to switch attention back and forth between the strategies (i.e., the available evidence would suggest that retrieval would not be a viable option on that trial). In sum, when considered in the context of the goal structures which surely are operative in this task domain, the assumption of nonparallel strategy execution appears to be at least as viable as is the instance theory assumption of independent, parallel strategy execution.

The NPS theory does not make any explicit assumptions about learning mechanisms, beyond the simple assumption that a single representation for each item is strengthened with practice. More detailed learning assumptions (e.g., assumptions about whether or not learning requires strategic effort to form new associations) are not necessary to make predictions about aspects of performance which will be the focus of this paper. It only needs to be assumed that learning of new associations (such as that between the problem and the answer) does occur with practice. The NPS theory does, however, make strong predictions about what types of strategy transitions will and will not be observed. According to the theory, information can be retrieved only if it is consistent with the focus goal. Once the focus goal shifts to the algorithm, direct retrieval of the answer will not occur. Thus, not only does the theory predict that parallel strategy execution from the outset of a trial will not occur, it also predicts that retrieval will not be initiated at any point during the execution of the algorithm. Data from children's arithmetic are generally consistent with this prediction. Siegler (1988) showed that children's strategy transitions in mental multiplication are typically direct transitions from the algorithm to memory retrieval. Subjects typically do not start the repeated addition algorithm for multiplication and then remember the answer before finishing the algorithm. Also, children to not appear to skip steps within the algorithm (Siegler, 1988; see also a discussion by Logan, 1988), another effect which is precluded by the NPS theory.

It is important to note, however, that the strategy transition predictions of the NPS theory do not preclude all types of within-algorithm transition effects. Consider the repeated addition algorithm for 4 x 7 discussed previously. If it is assumed that the knowledge of a given subject allows the addition steps which are required by this algorithm to be accomplished by direct retrieval from memory (i.e., 7 + 7 = 14; 14 + 7 = 21; 21 + 7 = 28 can all be accessed as facts in memory), then the theory does indeed predict that the only new association which can form as a consequence of practice is the direct association between the problem 4 x 7, and the answer. However, if the knowledge of the individual does not allow each step to be executed by single-step retrieval, then learning of new associations within the algorithm can occur. For example, a given subject might initially execute Step 2 of the example for 4 x 7 by decomposing 14 + 7 into 14 + 6 = 20, and then adding 1 to get 21. With practice, a new association directly linking 14 + 7 to 21 might be formed. This "mini-transition" to retrieval within the algorithm is allowed under the NPS theory because it is consistent with the goal to execute the second step of the algorithm. Data from a study by Carlson and Lundy (1992) represent a good empirical example of this type of within-algorithm strategy transition. They gave college students practice on a complex arithmetic algorithm which consisted of clearly identified subtasks. Their results for a consistent data condition (i.e., a condition in which problems with specific numbers were practiced repeatedly) showed that there was not a direct transition from algorithm to retrieval. Rather, subjects first learned to retrieve answers to the subtasks, and then learned to retrieve answers directly without retrieving answers to the subtasks.

The instance theory assumption that learning is an obligatory consequence of attention would presumably allow a variety of strategy transitions to take place which are not allowed under the NPS theory. In the repeated addition example, associations between early steps of the algorithm and the final answer, as well as associations which allow one or more steps of the algorithm to be skipped, might develop. Thus, although Logan (1988) implies that the association from the problem to the answer is the only one which develops with practice, this assumption does not fall out as a necessary consequence of his learning assumption.

It will be useful for later purposes to distinguish between two types of algorithms, which I will term <u>optimal</u> and <u>nonoptimal</u>, for which the NPS theory makes different strategy transition predictions. Optimal algorithms are those for which all of the algorithm goals are achieved by direct memory retrieval. An example is the repeated addition algorithm for multiplication discussed above in the case where the child can retrieve the answer to each addition step directly from memory. Nonoptimal algorithms are those for which one or more algorithm goals are achieved by a subgoaling process which essentially executes a "mini-algorithm". The Carlson and Lundy (1992) task is an example of a complex nonoptimal algorithm. The NPS theory predicts that only one strategy transition can take place when the algorithm is optimal at the onset of practice (the single-step transition to retrieval), whereas multiple transitions can take place when the algorithm is non-optimal at the onset of practice (mini-transitions within the algorithm as well as the transition to retrieval). In principle, it should be possible to identify the goal structure which is operating at the beginning of practice on a given task based on a priori consideration of the problem domain and of subject's knowledge, and/or on verbal protocol techniques (Ericsson & Simon, 1993). Given this preliminary task analysis, the NPS theory can make clear predictions about what types of transition effects will and will not be observable during the course of practice.

The evidence bearing on instance-based versus strength-based memory representation in this task domain is inconclusive. The classic evidence addressing this issue suggests that, in at least some learning contexts, strength is not enough. Strength must be supplemented, or replaced, by instance memory (for a review see Hintzman, 1976). Arguments from parsimony would lead one to reject strength assumptions altogether and opt for pure instance memory. Other factors, however, need to be considered before rejecting a strength approach. First, strength-based models (including the broad class of connectionist models) have been quite successful in a variety of domains to which instance-based memory models have not yet been applied (see Anderson, 1983; Campbell & Oliphant, 1992; MacKay, 1982). Thus it would be premature to conclude that strength-based models are not operative in some circumstances. Second, there are at least two important differences in the tasks being exploring here and the tasks which are typically used to support instance memory. First, the current tasks reflect the acquisition and repetition of new associative information. When the associations are new, it might be a very effective learning strategy to strengthen a single, initially very weak or non-existent, representation. In contrast, in the classic literature on instance memory, subjects are typically exposed during the learning phase to a large number of relatively familiar words, some of which are presented multiple times, and they are asked to study each word. Later, they are asked to recognize, recall, and or make frequency judgements about the words. In these tasks, strengths of any abstract or semantic representations of the words might be relatively close to asymptote initially, and thus attempts to establish unique episodes or instances representing each word presentation might be a more effective learning strategy. A second, related difference between the tasks is that the purpose of practice in the current tasks is very clear from the outset (i.e., learn to retrieve the answers quickly). This fact might lead to very homogeneous learning processes which would promote strengthening of the optimal type of representation for the task goal. In contrast, in the classic literature supporting instance memory, instructions in the learning phase were relatively vague (e.g., study the words). Learning strategies in this context may be much less homogeneous and perhaps much less likely to result in strengthening of a specific, goal oriented form of representation for each item.

An additional motivation for assuming strength-based memory in the NPS theory is that is it not clear how to integrate instance-based memory with the other assumptions of the theory. If strategies cannot be executed in parallel, then some index for monitoring the likelihood that the memory retrieval strategy will succeed is necessary for that strategy to be selected over the algorithm strategy. In a strength-based memory approach, a numerical value corresponding to memory strength provides a natural index. In an instance-based memory approach, in which instances are assumed to accrue independently of one another, it is not clear how such an index could be derived. One could stipulate some mechanism which "counts" the number of instances and feeds this information to a strategy selection mechanism. However, if a counting mechanism is allowed, then an instance theory becomes very similar, perhaps indistinguishably similar, to a strength theory. The argument here is not that instance memory could not conceivably be integrated with the other assumptions of the NPS theory, but rather that it is not presently clear how this could be achieved. In contrast, strength memory meshes very nicely with the other assumptions of the theory.

The empirical and conceptual arguments discussed above suggest that the assumptions of the NPS theory, although perhaps substantially oversimplified, are nevertheless at least as viable as are those of the instance theory. Convincing demonstration, however, that this theoretical perspective is preferable would require (a) specification of an explicit quantitative model which is consistent with the assumptions of the theory, (b) demonstration that this new model makes importantly different empirical predictions than does the instance theory, and (c) empirical verification of those predictions.

The Component Power Laws Model

In this section I will describe further specifications of the NPS theory in order to develop a testable quantitative model which I will refer to as the Component Power Laws (CPL) model. The model as developed here is intended to account for the following aspects of performance: (a) the speedup in mean RT and reduction in SD within each strategy as a function of practice, (b) the probability of using the retrieval strategy as a function of practice, and (c) the overall speedup in RT and reduction in SD as a function of practice.

Speedup and Reduction is SD within Each Strategy

In the instance theory, predictions for overall speedup in mean RT and reduction in SD can be derived mathematically from the basic quantitative assumptions of the model. The NPS theory does not lead to any obvious mathematical derivation of this type. It does, however, make one important prediction which allows what is known empirically about speedup and reduction in SD in certain contexts to be incorporated into a model. Specifically, the theory predicts that algorithm and retrieval strategies operate independently once they are selected. Once the algorithm is initiated, none of the attributes of the memory strategy (e.g., the strength of the memory representation for the problem) can influence performance. Similarly, once the retrieval strategy is selected, no attribute of the algorithm strategy influences memory retrieval. Thus, the selected strategy will be executed exactly as it would be in a task that does not exhibit the transition from algorithm to retrieval. It is therefore reasonable to assume that the functional characteristics of speedup and reduction in SD for the algorithm and retrieval strategies, when they occur in a task which exhibits a strategy transition, are the same as the functional characteristics of these strategies when they occur in isolation. The available evidence strongly indicates that both speedup and reduction in SD for these strategies in isolation follow power functions. First consider the retrieval strategy. Pirolli and Anderson (1985) demonstrated that speedup with practice follows the power function for verification of facts of the form, "The hippy kissed the debutante in the park." More recently, data reported by Rickard and Bourne (1994) demonstrate that speedup follows power functions in the domain of arithmetic fact retrieval. In this experiment, 24 subjects received extensive practice (90 repetitions over 4 sessions) on a set of 16 simple multiplication and division problems of the general form, "4 x 7 = __". The mean RTs and SDs for each block of practice are shown in Figure 2. There are no systematic visual or statistical deviations from log-log linearity for either measure.

Now consider the functional characteristics of speedup mean RT and reduction in SD for algorithms. An important factor here is the previously defined distinction between optimal and nonoptimal algorithms. An algorithm can be nonoptimal in an endless number of ways and it might prove very difficult to make strong conclusions based on empirical data which would hold for any conceivable non-optimal algorithm. By contrast, it is relatively straightforward to make empirically motivated conclusions about optimal algorithms. Recall that for optimal algorithms, each of the steps of the algorithm is executed by direct memory retrieval. Thus, optimal algorithms are essentially a string of successive fact retrieval events. It can be shown that if speedup and reduction in SD for a single fact retrieval follow power functions, then these variables will also follow power functions for a string of fact retrievals given that two very reasonable assumptions are made: (a) the overall RT for the algorithm is an additive function of the RT for each component retrieval, and (b) the rate parameters of the power functions for each of the component retrievals are approximately the same.

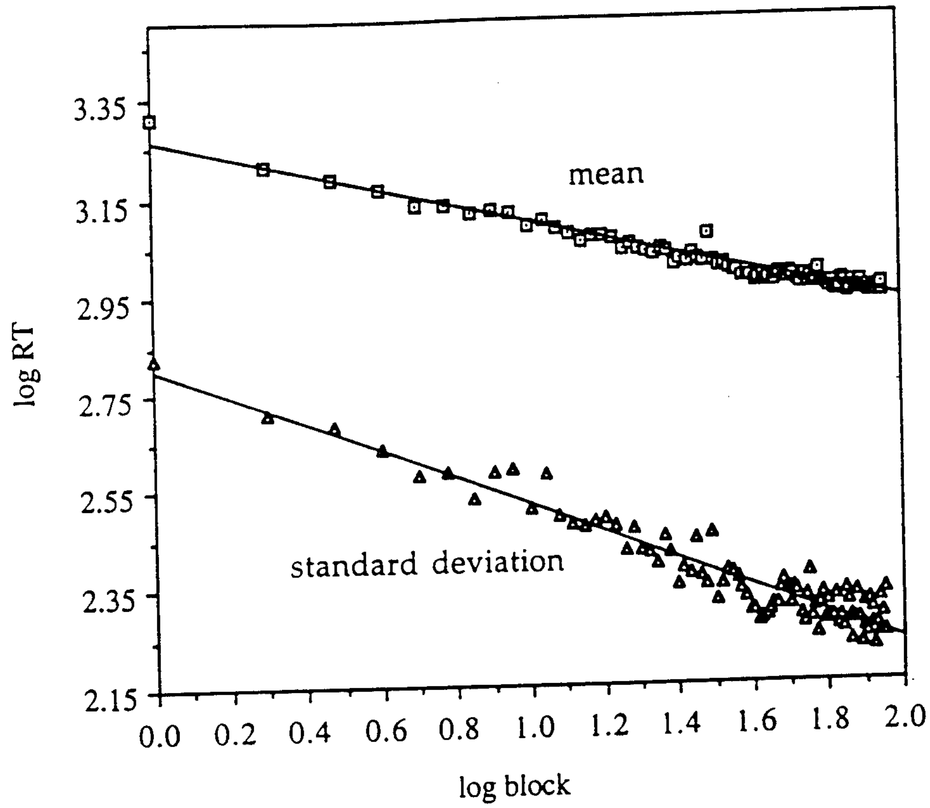


Figure 2. Log means and standard deviations of reaction times to verify true alphabet arithmetic equations as a function of the log of the number of presentations and the magnitude of the addend size. Lines represent fitted power functions constrained to have the same exponent for means and standard deviations. Reprinted from Logan (1988).

The first assumption is more or less required by Assumptions 2 and 3 of the NPS theory. The second assumption is robust and violations of it are unlikely to cause serious problems in most cases.

One additional factor which may result in speedup and reduction in SD of optimal algorithms is general speedup in algorithm execution not related to speedup in retrieval of component facts from memory. Carlson and Lundy (1992) showed that this type of general algorithm learning also follows the power function. Thus, given that Assumptions A and B above hold to a reasonable approximation across both fact retrieval speedup and general speedup, power function speedup should hold for the algorithm overall even when general speedup occurs.

The considerations above support the claim that speedup and reduction in SD follow power functions for fact retrieval and for optimal algorithms when these strategies occur in isolation. These facts can be used to make the same predictions when these strategies occur in the context of strategy transitions, given two technical assumptions about the strategy selection and execution. First, it is necessary to assume that in the majority of cases, once the retrieval subgoal is selected, an answer will subsequently be retrieved from memory and will be stated as the response. That is, it is strictly necessary to assume only infrequent use of the algorithm as a backup strategy when retrieval fails (Siegler, 1988). Second, strictly speaking, the power functions would only be expected to hold at the individual item level within each strategy when the transition is a step function, such that the algorithm is used for the first n trials, and retrieval is used for all of the remaining trials. Provided that the transitions for each item do not deviate in the extreme for a step function, however, deviations from power functions are likely to negligible.

Retrieval Probability as a Function of Practice

The strength-based memory assumption of the NPS theory provides a conceptual starting point for developing a quantitative model of the probability of retrieving the answer from memory as a function of practice on a given item. In the memory literature, the effects of strengthening on performance are often modeled using a simple strength-threshold model (Anderson, 1992; MacKay, 1982). Strength is assumed to increase gradually and monotonically as a function of practice. If strength is below a hypothetical threshold value, retrieval will not take place (i.e., the algorithm will be used), whereas if it is above threshold, retrieval will take place. This same approach is taken here. If there is no noise, or variance, in either the strengthening process or the threshold, then a strength-threshold model predicts a step-function transition, with the algorithm always used during the first n trials of practice (before strength reaches threshold), and retrieval taking place for all subsequent trials. This simple model may indeed account for strategy transitions at the item level in many cases. However, a more realistic model would also allow for some variance in either the strength or the threshold values as a result of variety of random influences (e.g., lapses of attention, interference from recently solved problems, other intrinsic noise factors). If these noise factors are assumed to be roughly normally distributed, then a reasonable mathematical model of the probability of retrieving as a function of practice is provided by the logistic function, which has the general form,

$$p = 1 - 1/(1 + e^{((N - a)/b)}), \tag{1}$$

where p is the probability of retrieving, N is the number of trials of practice, the parameter a corresponds to the trial number at which $p = .5$, and the parameter b determines the speed with which the transition takes place (it corresponds to the inverse of the instantaneous slope of p when $p = .5$). This function can exhibit a very fast transition from 0 to 1 which is essentially a step function when the value of b is small (around .1 or smaller). When b is larger, the function has a sigmoidal form symmetrical about the value of a.

Predictions for Overall RTs and SDs

The CPL predictions for overall speedup and reduction in standard deviation for a given item will reflect the combined influences of the component power functions for each strategy, and of the logistic strategy transition function specified in Equation 1. Process mixture equations (see Townsand & Ashby, 1984) require that the overall RT at any point during practice is given by,

$$RT = (1\text{-}p)*(RT_{alg}) + (p)*(RT_{ret}), \tag{2}$$

where p is the probability of retrieving determined by Equation 1, RT_{alg} is the RT for the algorithm, and RT_{ret} is the RT for retrieval. The overall variance is given by,

$$Var = (1\text{-}p)*(Var_{alg}) + p*(Var_{ret}) + p*(1 - p)*(RT_{alg} - RT_{ret})^2. \tag{3}$$

In Equations 2 and 3, the RTs and variances for the component strategies will be assumed to be two parameter power functions; asymptotes are ignored (assumed to be zero) for simplicity. In most cases this assumption can be made without introducing significant bias (Newell & Rosenbloom, 1981). Note that if the SD follows a power function, the variance must as well. Thus the empirical evidence that reduction in SD follows a power function within each strategy also implies power function reduction in variance in Equations 2 and 3.

Sample CPL RT and SD functions are shown in Figure 3 (a and b) in log-log coordinates. The curved lines in Figure 3 (a and b) are the overall CPL functions for the RT and SD, and the straight lines are component power functions for the algorithm and retrieval processes. In this example, I arbitrarily set the component functions for retrieval to have steeper slopes than those for the algorithm. For the RT, the deviation from linearity reflects a simple weighted averaging of the algorithm and retrieval processes at each point during practice. During roughly the first half of the transition to retrieval, the curve takes a concave downward form, and during the second half of the transition, it takes a concave upward form. For the SD, the relation between the overall function and the component functions is slightly more complex. Due the the third "bubble" term in Equation 3, the overall SD during roughly the first half of the transition period will always be larger than the algorithm SD.

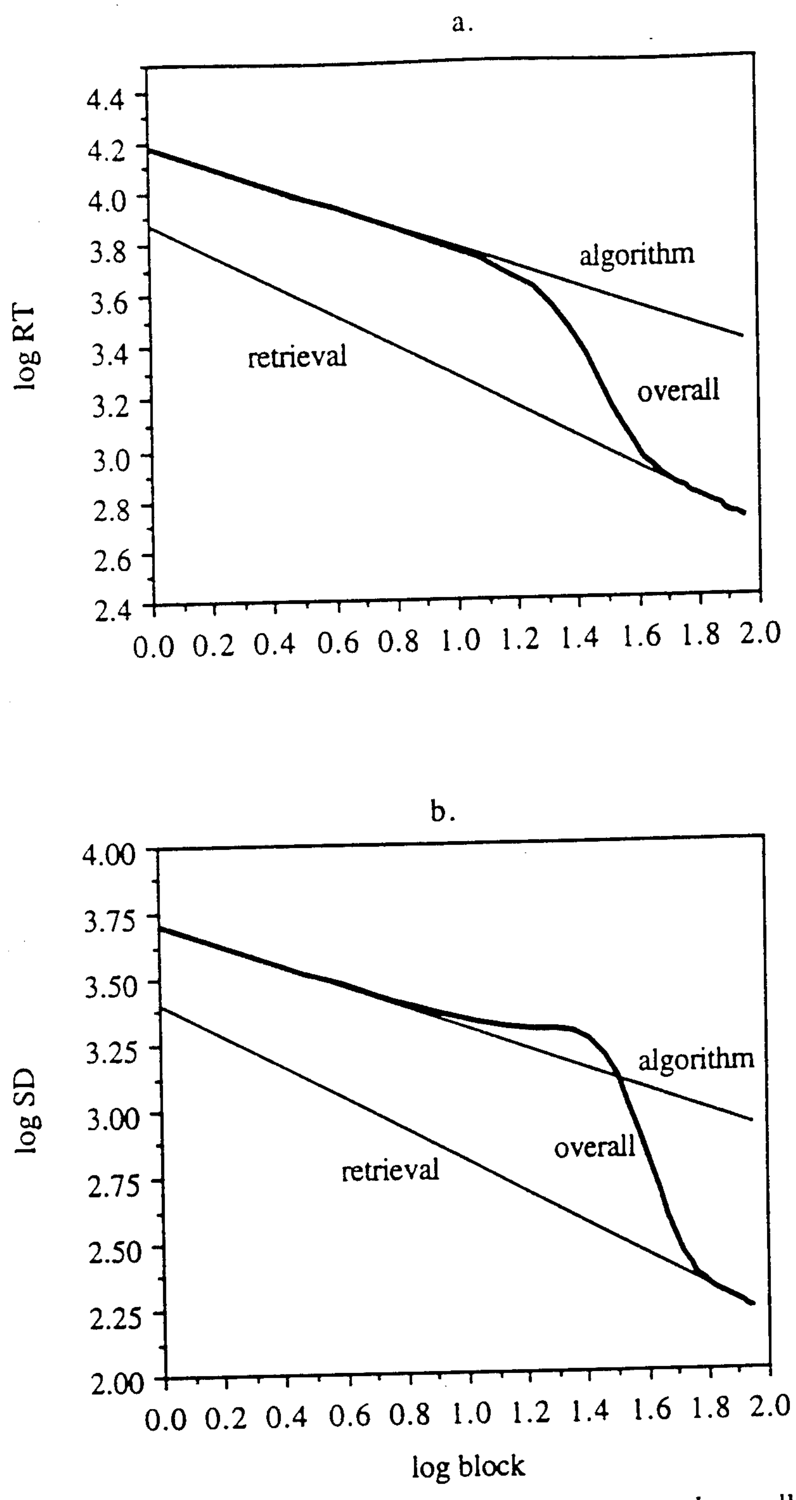


Figure 3. CPL predictions for sample component and overall log means (Panel a) and standard deviations (Panel b) of reaction times as a function log block. Thin lines represent predictions for the component strategies; thick lines represent predictions for the overall data.

Methodological Considerations in Fitting the CPL Model to Group Data

The CPL derivations above apply strictly only to a single item being solved by a single subject. In order test the model empirically, however, data will have to be collapsed across items and subjects. Several types of distortion of the predictions of the model can potentially be introduced by this data aggregation approach. First, even if power functions hold at the item level for a given data set, there is no mathematical guarantee that they will hold if data in raw form are collapsed across items and subjects. In practice, this potential source of bias is unlikely to present a problem. The data of Rickard and Bourne (1994), for example, were collapsed in raw form across items at the subject level, logged, and then averaged across subjects to yield the data shown in Figure 2. This problem can be eliminated entirely, however, if data are first logged at the individual item level, and then collapsed across items and subjects. This mathematically more appropriate approach will be used for all RT analyses in this dissertation. Analogously, the appropriate approach to aggregating SD's (used in these experiments) is to log SD's for each subject for each block of practice, and then to average these values across subjects.

In the experiments described below a strategy probing technique is used to separate trials into cases in which the algorithm was used and into cases in which retrieval was used. Evaluation of whether the power function holds within each strategy is based on these strategy probes. An intrinsic complication of separating across strategies in this way is that due to differing rates with which the transition to retrieval takes place across items and subjects, the RTs and SDs within each strategy at any point during practice will represent constantly changing subsets of the items and subjects. The possibility that the number of trials necessary to make the transition to retrieval is correlated with RTs or SDs across items and/or across subjects represents an additional possible source of bias in the RT and SD analyses for the individual strategies. Consider for example subject *a* whose RTs for a given strategy (say, the algorithm) follow a power function with a small intercept value (i.e., fast RTs) and who makes the transition to retrieval only after many exposures to each item, and subject *b* whose RTs for the algorithm follow a power function with a large intercept and who makes the transition to retrieval very quickly. The average of RTs in this case will not follow a power function. There is no a priori way to correct for these possible distortions. Rather, if deviations from log-log linearity are observed within a strategy, empirical checks for this possible correlation are needed before concluding that the predictions of the CPL model do not hold. Note that this distortion effect can occur only when data are analyzed separately by strategy; any distortions which may be observed in the overall RT or SD data cannot be attributed to this effect because each item and each subject is reflected in these data throughout practice.

A second potential source of distortion in SD data (but not in the RT data) for the component strategies (but not for the overall data) arises from the approach discussed above of logging the SDs at the block level for each subject and then averaging these logged values across subjects. Take the algorithm SD estimates as an example (although the same problem exists for the retrieval SDs). As practice progresses, the number of observations constituting the algorithm SD estimates for each subject will steadily decrease, because the algorithm is used less and less often with practice. Thus there will be more noise associated with the estimates of the SDs later in practice than is associated with the estimates early during practice. Because SDs are bounded at the lower end by zero, the increasing noise will manifest primarily as occasional very large SD estimates. Under these conditions, taking the logs of the data before averaging across subjects will bias the obtained means of the SDs to be smaller when these SDs are based on a small number of observations than when they are based on a large number of observations. Simulations using data and item transition patterns closely matching those obtained in the current experiment, however, showed that these deviations from linearity are essentially nonexistent as long as data are analyzed only for blocks of practice on which most or all of the subjects contribute an SD estimate. This approach will be taken in the current experiments.

Finally, it is easy to see that significant distortions in the transition curves (the probability of retrieving as determined by Equation 3) can potentially occur when data are collapsed across items and subjects. Consider for example the (unlikely) possibility of a bimodal distribution in the transition data, such that the transition to retrieval occurs for one-half of the items very early during practice, and for the remaining items much later. The empirical transition curve collapsed across items in this case would have a steep upward slope initially, would then level off in between the two groups of items at a value of p=.5, and would then slope upward sharply as the transition occurs for the second group of items. This transition function clearly differs from the predictions of the logistic model. The best solution to this problem is to fit logistic curves separately for each item for each subject, and then to predict the overall empirical transition curve based on the average of the theoretical curves for each item. This approach will be discussed in more detail in the Methods section of Chapter 2.

A CPL Account of Two Empirical Results in the Literature

Two results in the skill acquisition literature provide preliminary support for the CPL model. First, the complex arithmetic algorithm of Carlson and Lundy (1992) discussed earlier included two distinct data conditions. In the consistent data condition, the same problems were presented repeatedly throughout practice. In the varied data condition, problems were varied from block to block such that any given combination of numbers was presented as a problem only once during practice. The CPL model predicts that, in the consistent data condition, a transition to retrieval will occur, but that in the varied data condition, speedup in RT and reduction in SD will reflect only increases in efficiency of algorithm execution. Thus deviations from power functions should be observable in the consistent data condition, but should not be observable in the varied data condition. To investigate this possibility, I replotted the data of Carlson and Lundy (1992) as shown in Figure 4. Least squares second-order polynomial

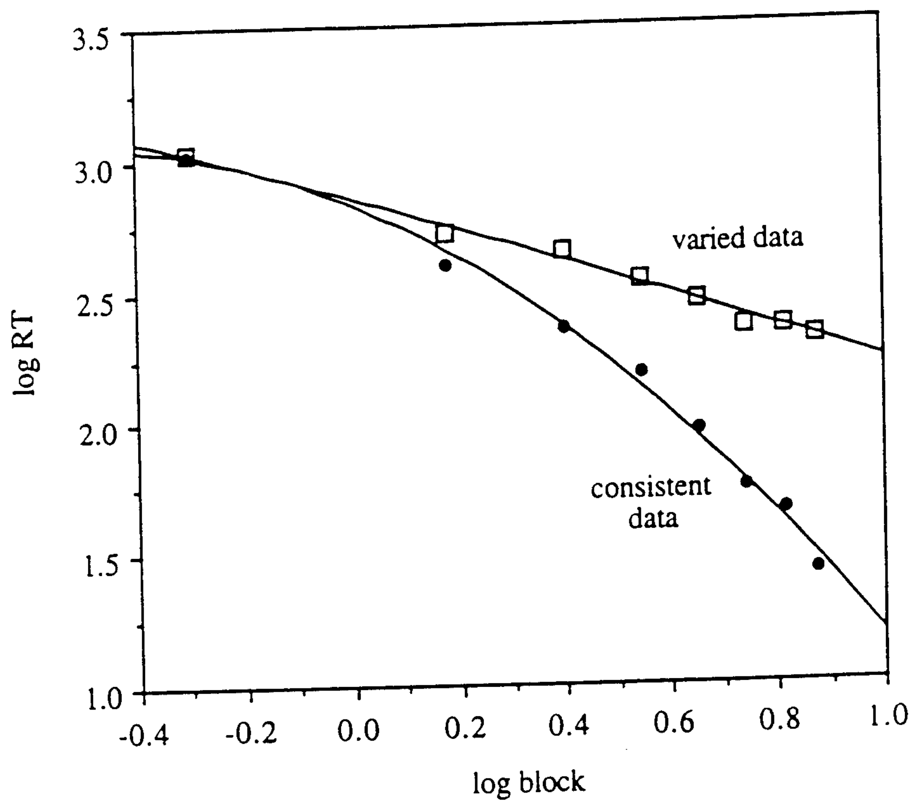


Figure 4. Log means of reaction times to perform the complex arithmetic task of Carlson and Lundy (1993) plotted as a function of task type (varied versus consistent data) and the log of the number of presentations. Fitted lines represent second-order least squares regression equations.

regression equations were fit to the data from both conditions to highlight the functional differences in speedup. In the varied data condition, the power function fits nicely with no deviations from linearity evident. In the consistent data condition, however, concave downward nonlinearity is clearly visible. These effects in their general form are consistent with the predictions of the CPL model.

The CPL model can also in principle account for each of the results of Logan's (1988, Experiment 5) alphabet arithmetic experiment, including those results which are are not easily accounted for by the instance theory. Recall that in that experiment there were concave-downward deviations in both the RT and SD data which became larger with increasing addend size (see Figure 1). Also, these deviations were more extreme for the SD data than for the RT data. Protocols collected at the end of each session (Logan, 1988) indicated that the transition to retrieval was only about 70% complete at the end of the last session. Under these conditions, the CPL mode predicts a concave-downward deviation in both the RT and SD data. Because the transition to retrieval was not complete, the lack of any clear concave upward deviations following the concave downward deviation is also consistent with the model. The fact that the deviations increase with increasing addend size is also consistent with the model because the more time consuming the algorithm, the greater the "distance" that has to be traveled between the algorithm and retrieval power functions (assuming that the retrieval functions are unrelated to addend size) and thus the greater the deviations from linearity evident in the figures. Finally, the fact that the deviations from linearity are larger for the SD data is also consistent with the model due to the influence of the "bubble term" in Equation 3 (see the example CPL functions for the SD in Figure 3). In the following chapters, two experiments designed to provide more definitive tests of the CPL model are described.

CHAPTER 2

EXPERIMENT 1

A pseudo-arithmetic task which I will refer to as *pound arithmetic* was developed as an initial test of the CPL model. Two types of pound arithmetic problems were constructed using a generic arithmetic series in which the third element of the series is the difference between the first two elements, plus 1, added to the the second element. For example, the third element of the specific number sequence 9, 15, ?, is [(15 - 9) + 1] + 15 = 22. In Type 1 problems, the third element of the series was unknown (e.g., 9 # 15 = __). In Type 2 problems, the second element of the series was unknown (e.g., 9 # __ = 22). Problems were presented in a traditional arithmetic format (as in the examples above) with a blank holding the place of the missing element, and with the # symbol used to hold the place of the arithmetic symbol. Subjects were taught a three-step algorithm, as shown above, for solving Type 1 problems, and a related four-step algorithm for solving Type 2 problems.

This task appears to be well suited for testing the CPL model for several reasons. First, it is very likely that the algorithms for both problem types are optimal or nearly so for most college students. Consider the example 9 # 15 = __ given above. The first step is to subtract 9 from 15, the second step is to add 1 to the result from Step 1 (6 + 1 = ?), and the final step is to add 7 to 15. Each of these arithmetic operations is probably executable as direct fact retrieval for most college students (see Ashcraft, 1992, for a review of the evidence that, for adults, simple arithmetic reflects fact retrieval processes). Thus, the CPL model predicts that only two strategies will be used; execution of the algorithm and direct retrieval of the answer from memory. A second motivation for using the pound arithmetic task is that it is very similar to standard arithmetic, and thus the findings of this experiment should generalize to acquisition of these skills.

There are also two reasons why the pound arithmetic task appears to be a good candidate for exhibiting nonparallel strategy execution as predicted by the NPS theory. First, the algorithm involves basic but nontrivial arithmetic fact retrieval processes as well switching among the arithmetic operations of addition, subtraction, and division. These facts, combined with the fact that subjects must choose and discriminate between two algorithms (those for Type 1 and Type 2 problems) on each trial, suggests that attentional demands of the algorithm will be high. These high attentional demands might force a choice between the algorithm and direct retrieval strategies at the onset of each trial. Second, the direct retrieval of the answer and execution of the component steps of the algorithm (which themselves are arithmetic fact retrieval processes) are intuitively similar cognitive events and thus may require the same cognitive "modules". If this speculation is correct, and if the involved module can effectively execute only one retrieval process at a time, then interference effects may preclude parallel execution of the two strategies. In combination, these points suggest that, if the empirical predictions of the CPL model hold for *any* task, they should hold for pound arithmetic. Thus, this experiment can be seen as providing a test of the CPL model in a maximally "friendly" task environment. A task which has very different characteristics will be explored in Experiment 2.

Yet another motivation for investigating the pound arithmetic task is that it allows for a new test of a recent *identical elements* model of the memory structure for basic arithmetic facts (Rickard, Healy, & Bourne, in press; Rickard & Bourne, 1994). The identical elements model, which was developed to account for the structure of memory for basic multiplication and division facts in adults, assumes a single and functionally distinct representation in memory corresponding to each unique combination of the two numbers (ignoring order) which constitute a problem (e.g., 4 and 7), the number that is the answer (e.g., 28), and the arithmetic operation formally required to produce the answer (e.g., multiply). The model assumes distinct perceptual, cognitive, and motor stages of arithmetic fact retrieval (see also McCloskey, Caramazza, & Basili, 1985), and it applies to the structure of knowledge as represented within the cognitive stage. Problems that have exactly the same elements will access the same memory representation within the cognitive stage, despite any perceptual differences, such as format or modality of presentation. For example, multiplication problems that differ only in operand order, such as 3 x 8 and 8 x 3, will access the same representation. In contrast, problems that differ with respect to even one element will access completely different representations. For example, complementary problems from two operations (e.g., 4 x 7 = _ and 4 x _ = 28), and complementary problems within a non-commutative operation (e.g., 28 = _ x 4 and 28 = _ x 7), access completely different representations. Data from arithmetic studies in which adult subjects were practiced extensively on a set of basic multiplication and division problems, and then tested on various altered versions of these problems, confirm the basic predictions of the model. For example, practice on one operand order in multiplication (e.g., 4 x 7 = __) transferred completely to the reverse order (e.g., 7 x 4 = __), once the perceptual advantage for the practiced order was factored out (Rickard & Bourne, 1994). In contrast, there was no transfer of learning to test problems that represent (a) a change in operation (e.g., 4 x 7 = __ during practice, 4 x _ = 28 at test), or (b) a "reversal" of operands for a noncommutative operation (e.g., 4 x _ = 28 during practice, 7 x _ = 28 during test).

In the current experiment, subjects were first practiced extensively on a set of Type 1 and Type 2 problems, and were then tested on the exact problems seen during test (no-change problems), on type change problems (i.e., a Type 1 problem seen during practice was presented as a Type 2 problem), and on new problems not seen during practice. The identical elements model predicts that practice should result in a transition to retrieval only for no-change problems, despite the strong similarity of each type change problem to its corresponding no-change problem (e.g., 9 # 15 = __ and 9 # __ = 22). Both type change and new problems at test will be solved by way of the algorithm. Performance in these conditions should thus be roughly equivalent and also much slower than performance in the no-change condition. Note, however, that the identical elements model does not rule out the possibility of somewhat improved performance in the type change and new problems conditions at test, relative to performance at the beginning of practice, because the possibility of general speedup in algorithm execution time is not inconsistent with the model.

Method

Subjects

Twenty-one subjects from an introductory psychology course participated in the experiment for credit. Two of these subjects were replaced because they failed to attend all of the practice sessions. An additional subject's strategy probing data revealed that no transition to retrieval occurred during the course of practice. This subject was also replaced to yield a total of 18 subjects who attended all sessions and showed a transition to retrieval with practice. The data from the single *nontransition subject* were preserved, however, for separate analysis. All subjects were tested on Zenith Data Systems personal computers, programmed with the Micro Experimental Language (MEL) software (Schneider, 1988).

Apparatus and Materials

Three subsets of 6 pound arithmetic problems were constructed. Within each subset, there was one problem with each of six left-hand numbers (3 through 8), and there was at most one problem with each of nine middle numbers (11 through 19), and at most one problem with each of 18 right-hand numbers (18 through 35). Three master sets of 12 problems were then created, one from each of the two-way combinations of the three subsets of six. Six experimental problem sets were then created, two from each master set (see Appendix A). One of the two problem sets created from each master set had one subset of six problems written as Type 1 problems, and the other subset written as Type 2 problems. The other problem set reversed the problem types (e.g., a Type 1 problem became a Type 2 problem). Each subject solved problems from only one experimental problem set during practice. Thus, each subject saw 12 problems during practice, six Type 1 problems, and six Type 2 problems. Four subjects showing a transition to retrieval were practiced on each of the six of the problem sets. During subsequent immediate and delayed transfer tests, all subjects solved all 18 problems presented as both Type 1 problems and Type 2 problems.

Procedure

The experiment lasted for 6 sessions, the first three on Monday, Wednesday, and Friday of one week, two additional sessions on Monday and Wednesday of the following week, and a final session on the Wednesday 6 weeks after the fifth session. Each session lasted 30-45 min. Subjects were tested in groups of up to 4. At the beginning of the first session, the subjects were given an example sheet describing the algorithm and an example problem worked out step by step for each problem type. The experimenter worked these example problems on a blackboard, with the subjects following along using the example sheet. The subjects were then given six problems (3 Type 1 problems and 3 Type 2 problems) to work independently using paper and pencil (these problems were different than those used in the main experiment). When the subjects completed the problems, the experimenter checked the results for accuracy and made corrections where necessary, making it clear to the subject what the errors were, and what they should do differently to correct them. From this point on, subjects performed the task independently at their own computer without the benefit of pencil or paper, although subjects were allowed to take the algorithm sheet with the example problems with them to the computers. For the remainder of the first session, subjects performed 9 blocks of problems using the computer, where each block was one exposure to each of the 12 problems in the subject's practice set. Problems were presented one at a time in the middle of the screen. Subjects entered the two digit answer using a number keypad on the right-hand side of the computer keyboard. Subjects were instructed to work as fast as possible while being accurate. They were told that they could rest briefly between blocks of problems. Latencies were collected from the onset of the problem to the pressing of the first digit of the answer (the initiate RT), and from the pressing of the first digit of the answer to the pressing of the second digit of the answer. The subject's answer for each problem was also collected.

Following one-third of the problems, subjects were probed for the strategy that they used. On these trials, a screen with three options was displayed below the problem about 1 sec. after they pressed the second digit of the answer. The options instructed the subject to press a special key marked "A" if they used the algorithm that they were taught to solve the problem, to press a key marked "R" if they retrieved the answer directly from memory, and to press a key marked "O" if they used some other strategy that did not correspond closely to either of the other options. Across every set of three consecutive blocks, each problem was probed once. Four problems were probed per block. Problems probed on each block were randomly determined, subject to the preceding constraint. The subject's strategy response, as well as the latency from the onset of the strategy options screen to the pressing of the strategy response, were both collected.

The second, third, fourth and fifth sessions consisted of 15, 21, 24, and 21 blocks of problems, respectively, presented on the computers as described previously. A transfer test was given immediately after the fifth session, which consisted of 3 blocks, each block consisting of one exposure to each of the 18 problems shown as both types, for a total of 36 problems per block. During the test, subjects were probed after every problem, in the manner described above. The delayed transfer test was exactly the same as the immediate transfer test, with the exception that no practice was given prior to the delayed test.

Results and Discussion

Practice: General

In the following analyses I will exclude from the data the single nontransition subject, focusing on the 18 subjects who reported a strategy transition with practice. Results for the nontransition subject will be discussed separately at a later point. Results for Type 1 and 2 problems were remarkably similar. There were no reliable problem type differences in terms of error rate, rate of transition to retrieval, or RTs. Thus, all analyses reported below were collapsed across this variable. Overall error rates were .109, .065, .055, .029, and .019 in Sessions 1, 2, 3, 4, and 5, respectively. A within-subjects analysis of variance (ANOVA) showed that the decrease in error rate across sessions was reliable, $F(4, 17) = 11.1$, $p < .001$. All subsequent analyses were performed on data from correctly solved problems.

The strategy probing results are shown in Figure 5, collapsed across subjects and problems, and across consecutive three block sequences across which each problem was probed once. Practice was successful in creating a transition to retrieval. By about block 60, retrieval was the reported strategy on nearly all trials. There were relatively few "other" responses, a result which supports the CPL prediction that pure algorithm and retrieval strategies are the only strategies which are used in this task.

Practice: Instance Theory Fits

To test the predictions of the instance theory regarding strategy transitions I fit a one-parameter power function of the form, $p = N^{-c}$, to the proportion of trials on

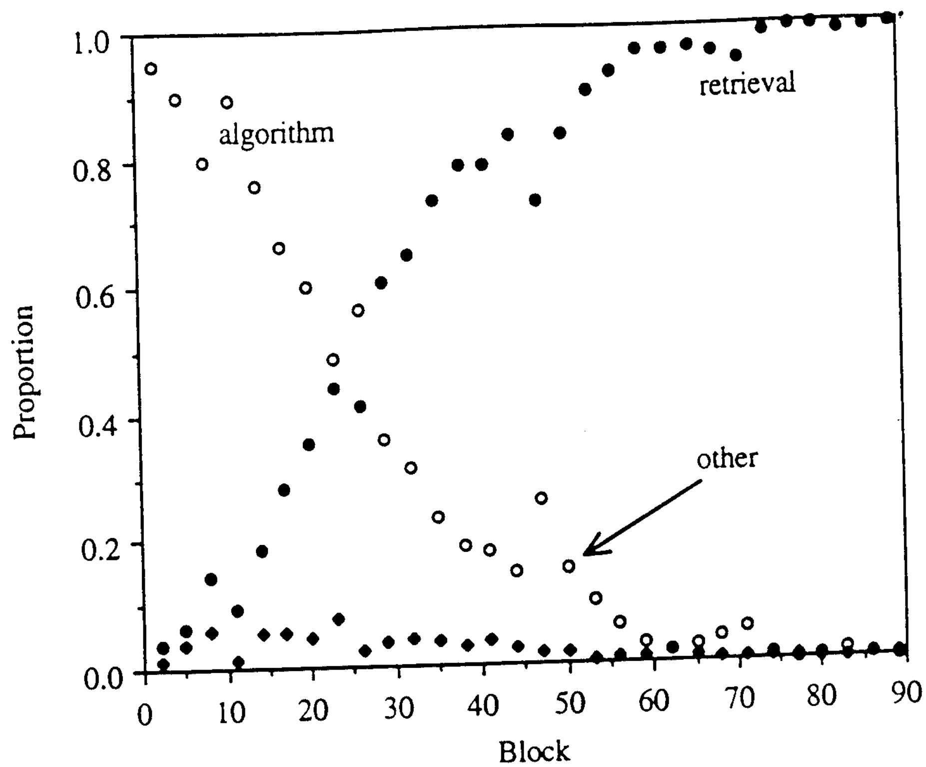


Figure 5. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *algorithm*, *retrieval*, and *other* strategies were reported as a function of block (Experiment 1)

which the algorithm was the reported strategy.[1] In this equation, P represents the predicted proportion, N is the number of practice trials (or number of blocks of practice), and a is a rate parameter. As discussed earlier, the instance theory does not strictly predict a power function reduction in the proportion of algorithm trials. Nevertheless, the function that it does predict will correlate highly with the power function, and any significant deviation from the power function can be taken as preliminary evidence against the theory. As shown in Figure 6, the fit was poor,

[1]Parameters a and b were not estimated because the most reasonable value of a, the asymptotic algorithm probability, is 0, and b must take a must take a value of 1 if the algorithm is the only available strategy initially.

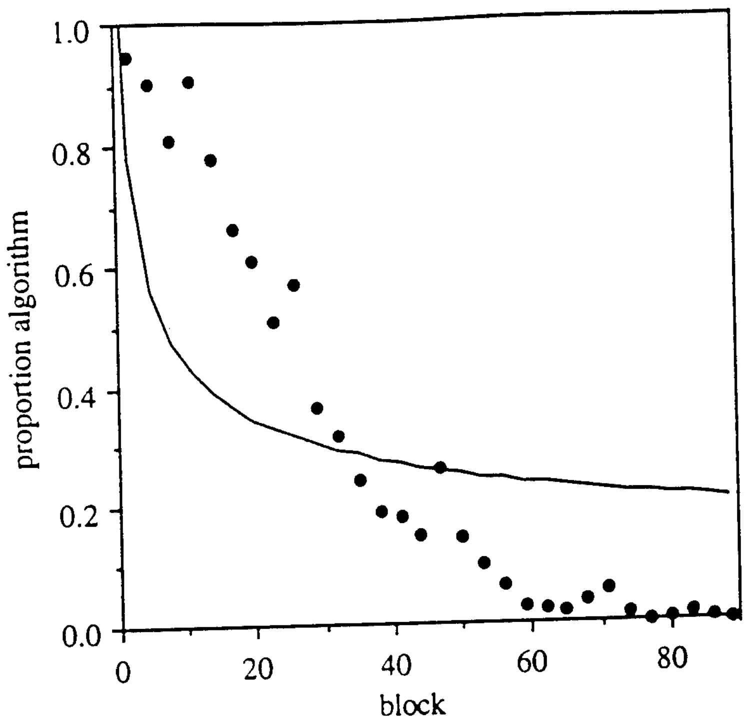


Figure 6. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *algorithm* strategy was reported as a unction of block. Fitted line is a single-parameter power function as discussed in the text, which is a close approximation to the predictions of the instance theory (Experiment 1).

yielding an r^2 of only .53 and exhibiting systematic visual deviations from the data.

Figure 7 (a and b) shows the log RT and log SD averaged across subjects and problems, plotted as a function of log block. Also shown in these figures are the best fitting power functions predicted by the instance theory (r^2 values were .93 and .88 for the RT and SD fits, respectively). Note the systematic deviations of the observed from the predicted values for both the RT and SD. In the early to middle stages of practice, the predicted values substantially underestimate the actual values (by a full second or more), and by the end of practice, the predicted values underestimate the observed values (again by about a second). Also, as with Logan's (1988) alphabet arithmetic data, the deviations from linearity are more extreme for the SDs than for the RTs.

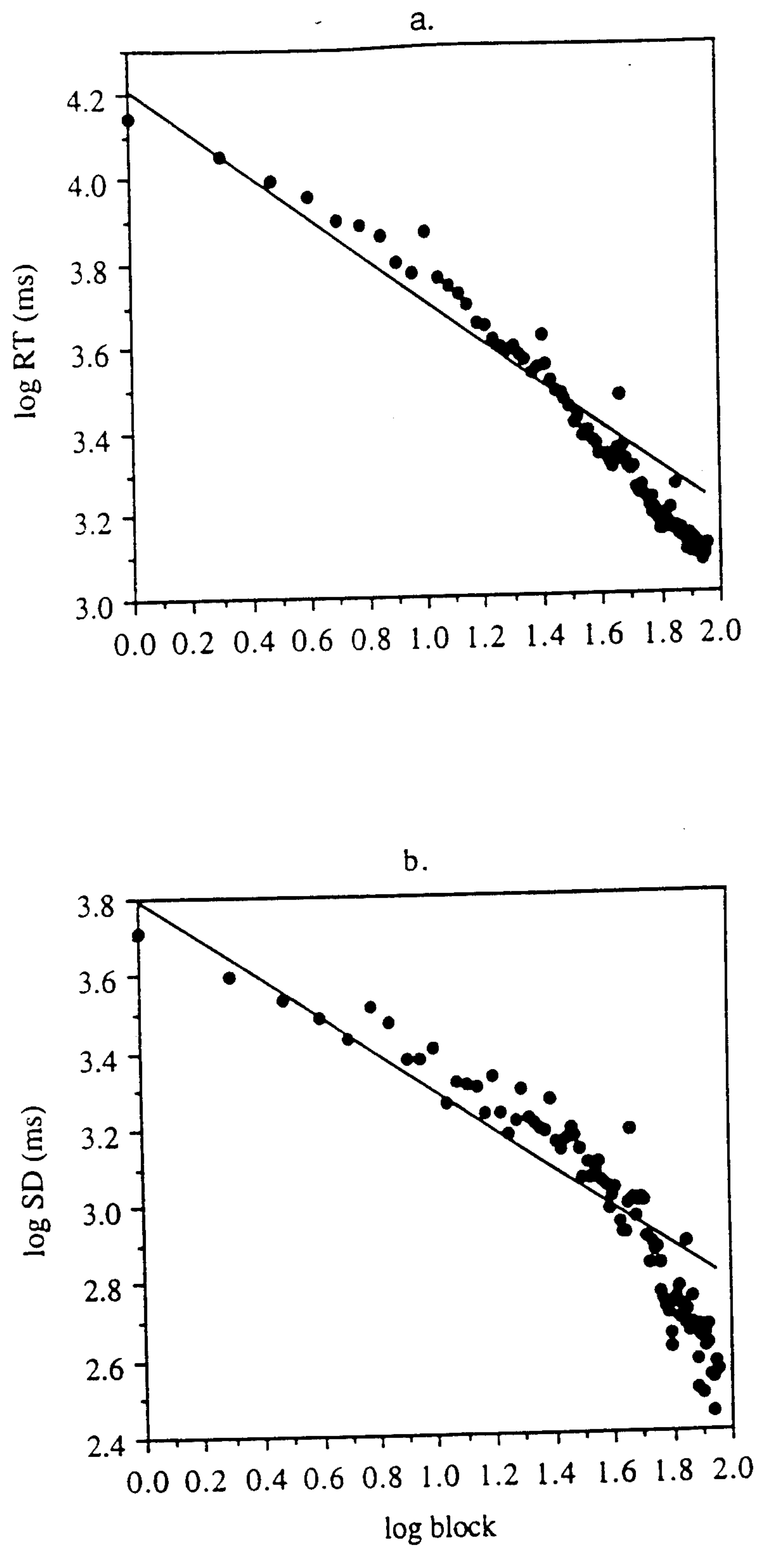


Figure 7. Log means (Panel a) and standard deviations (Panel b) of reactions times (collapsed across problems and subjects) plotted as a function log block. Fitted lines represent best fits of the instance theory (Experiment 1).

The instance theory predicts that the rate parameters for the RTs and SDs are the same. This prediction was tested by fitting 3 parameter power functions, which included a parameter for the asymptote, separately to each subject's RT and SD data. Sixteen of the eighteen subjects showed steeper rate estimates for the RT (mean = -.505) than for the SD (mean = -.435), a difference which was reliable by a binomial sign test ($p < .01$).

Practice: CPL Fits to the Strategy Transition Data

The CPL model prediction that the probability of using the retrieval strategy at the item level follows a logistic function of practice was tested by fitting a logistic curve to each item separately for each subject. In this analysis, strategy probe results were coded with a value of zero if *algorithm* or *other* response was given, and with a value of one if a *retrieval* response was given. Thus, for each item, the data consisted of up to 30 zero's and one's across 90 blocks of practice (the value was not always 30 because trials on which the response was incorrect were eliminated from the analysis). Ninety of the 216 items (41%) showed a pure step function transition, and thus the logistic fits to these items were essentially error free. The parameter a, which indicates the point at which the theoretical midpoint of the transition to retrieval occurred, was approximately normally distributed with a mean value of 24.6 and an SD of 12.63. The remaining items did not exhibit step functions and evaluation of the quality of the logistic fits to these items was accomplished using a somewhat more complex standardization approach. First, the estimate of the parameter a as determined by the logistic fit to each item was subtracted from the value of the block variable for that item. This transformation centered the data over the block variable in such a way that the predicted midpoint of the transition for each item occurred at block 0. Second, the block variable for each item was divided by a quantity (again determined by the logistic fit to each item) designed to transform the steepness of rate of the transition to be equivalent for each item. More specifically, the block variable for each item was adjusted so that the predicted probability of retrieving was .05 at Block -1, and .95 at Block 1. This transformation guarantees that the same logistic function will provide the best fit to each item; namely, a logistic function with values of 0 and .0526 for the parameters *a* and *b*, respectively. The transformation thus allows the data to be averaged across all items, while still maintaining the prediction that the logistic function will hold. The results are shown in Figure 8. The logistic function yielded an r^2 of .98, and exhibited no major systematic deviations from the observed data.

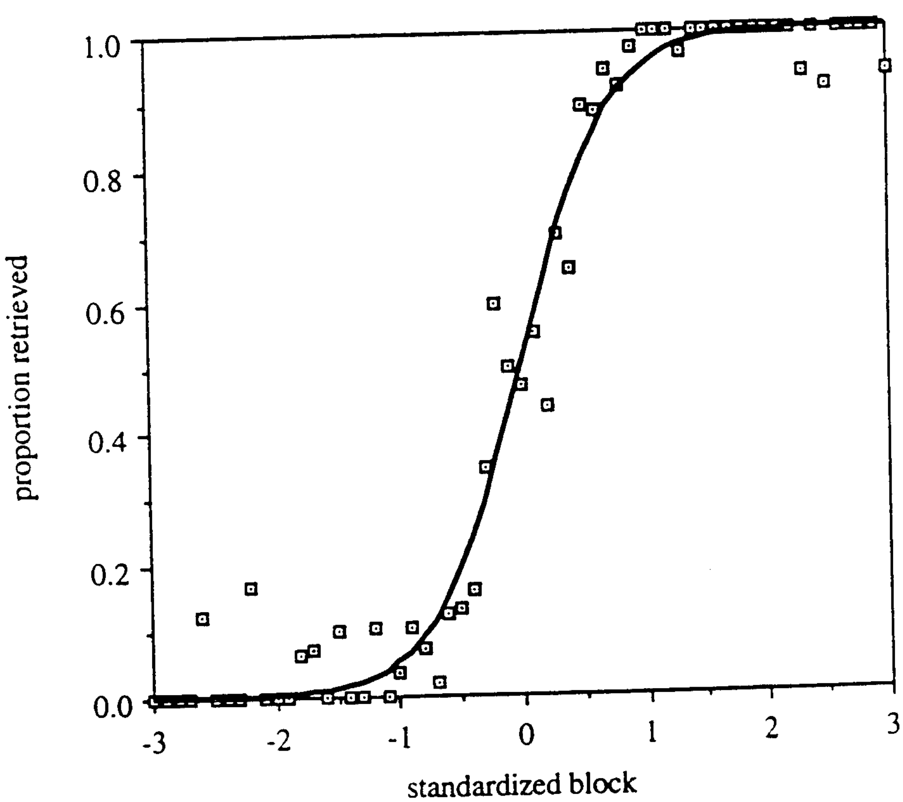


Figure 8. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *retrieval* strategy was reported as a function of standardized block. Problems for which the transition was a pure step function were excluded. Fitted line is the best fitting logistic function (Experiment 1).

Given that the logistic function provides a good model of the transition to retrieval at the item level, it is worthwhile to examine in some detail the estimates of the parameter *a* and *b* across items and subjects. Frequency distributions of *a* and of *spread* (a value derived from *b* which corresponds to the interval of blocks between predicted retrieval probabilities of .05 and .95 in the raw untransformed data) for all 216 items is shown in Figures 9 (a and b). The transition midpoint occurred most frequently around block 28, and dropped of rapidly in both directions with a slight right skew. The frequency distribution for the spread (Figure 9 b) has a very different character. The transition to retrieval occurred most often over an interval of 2 to 5 blocks, and occurred relatively infrequently over each of the larger intervals. Indeed, 70% of the transitions occurred over a 5 block or smaller interval. The combined facts that the midpoint of the transition has a roughly bell shaped distribution centered at around block 28, and that the majority of transitions took place very quickly (within about 5 blocks), provides converging evidence, along with the standardized logistic fits discussed previously, that the logistic function provides a good quantitative model of the strategy transition effects for this task.

The minimum, range, mean, and SD of the *a* estimates are shown separately for each subject in Table 1. Several statistical models of the distribution of the estimates of *a* were considered in an effort to shed more light on the learning mechanisms which underlie the transition to retrieval in this task. First, note that a simple hypothesis that a single distribution holds across subjects can be dismissed based on the wide variation in the summary statistics as shown in Table 1. It is still possible, however, that a single class of distributions, such as uniform or normal distributions, holds for all subjects but with different parameter values for each subject. One possibility is that transition midpoints are uniformly distributed with a minimum value on the second block of practice (the first block on which retrieval logically can occur) and a maximum

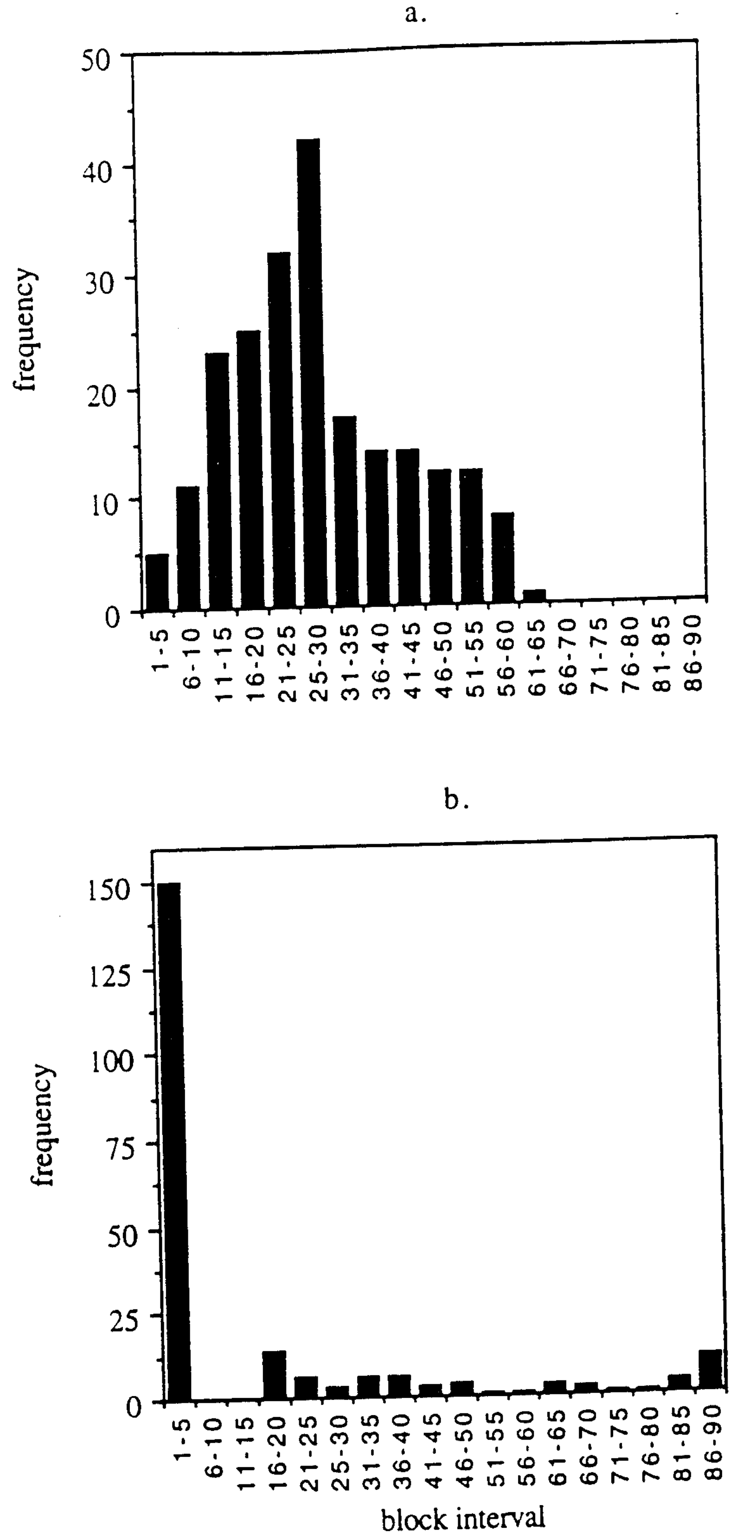


Figure 9. Frequency distributions for the parameter *a* (Panel a) and or the *spread* (Panel b) associated with the logistic fits for all items (Experiment 1).

Table 1

Summary Statistics on Item Transition Midpoints in Experiment 1.

Subject	min	range	mean	SD
1	7.93	30.12	25.9	9.76
2	23.49	16.5	31.1	4.68
3	13.7	46.2	34.5	15.8
4	5.5	22.5	20.1	7.4
5	7.2	48.6	29.7	14.8
6	14	19.9	24.5	7.8
7	7.7	44.8	35	14.4
8	7.56	50.8	34.6	18.2
9	5.5	51.4	35.8	16.3
10	4.5	21.5	15.9	7.38
11	6.5	50.9	43.9	15.39
12	15.7	10.3	22	3.1
13	5.5	42	25.7	14.4
14	3.1	24.9	16.4	7.4
15	15.5	26.8	22.6	7.27
16	24	33	38.8	12.3
17	15.7	20.3	26.7	7.3
18	29.7	35.6	42	10.9

value corresponding to the last observed transition midpoint for each subject. This model is generally consistent with a candidate hypothesis that subjects focus limited resources for learning sequentially on one item at a time starting at the beginning of practice. For example, subjects might adopt a strategy of selecting one item at a time to rehearse intermittently in the course of solving other items, and might move on to rehearse a new item only after the current item is sufficiently well learned to be retrieved reliably from long-term memory. If such a learning process accounts for all learning in the task, then it would predict a roughly uniform distribution of transition midpoints throughout practice for each subject. The rate with which learning would occur, however, might vary from subject to subject due to basic subject differences in speed of memorizing. This hypothesis was tested by computing z-scores for each subject, based on theoretical means and SDs associated with uniform distributions with the appropriate ranges for each subject, and collapsing these standardized data into a single distribution, shown in Figure 10a. If the hypothesis above is correct, then the data in the figure should have a uniform distribution with a shape identified by the dashed lines in the figure. Obviously the model is incorrect, and a chi-square test of goodness-of-fit for the uniform distribution confirms this fact, $X^2\ (7, 216) = 33.2$, $p < .01$.[2]

The failure of the simple uniform model may be due solely to the fact there was typically a large gap for many subjects between the beginning of practice and the earliest block of practice on which a retrieval midpoint occurred (see Table 1). It is possible that the distribution of retrieval midpoints is uniform for each subject, but with an arbitrary minimum value rather than a minimum value of 2 as assumed in the above model. This model would be consistent with the possibility that subjects use a sequential learning strategy as discussed above, but that they only adopt such a strategy after a number of blocks of practice. A third distributional hypothesis is that the retrieval midpoints are roughly normally distributed for each subject (albeit with different means and SDs). Normally distributed retrieval midpoints would be more consistent with the possibility of a single type of generic strengthening process which takes place independently for each item. According to such a model, strength for each item approaches the retrieval threshold at roughly the same rate on average. However,

[2]This and all subsequent chi-square tests for the uniform distribution were performed on frequency counts derived by dividing the data into 10 categories with equal intervals from a minimum z-score value of -1.73 to a maximum of 1.73, as shown in Figure 10. The values correspond to the theoretical minimum and maximum, respectively, of z-score values given that the data conform to a uniform distribution. Data outside of this range were not included in the observed frequency counts of any category. The theoretically expected counts, however, included all data.

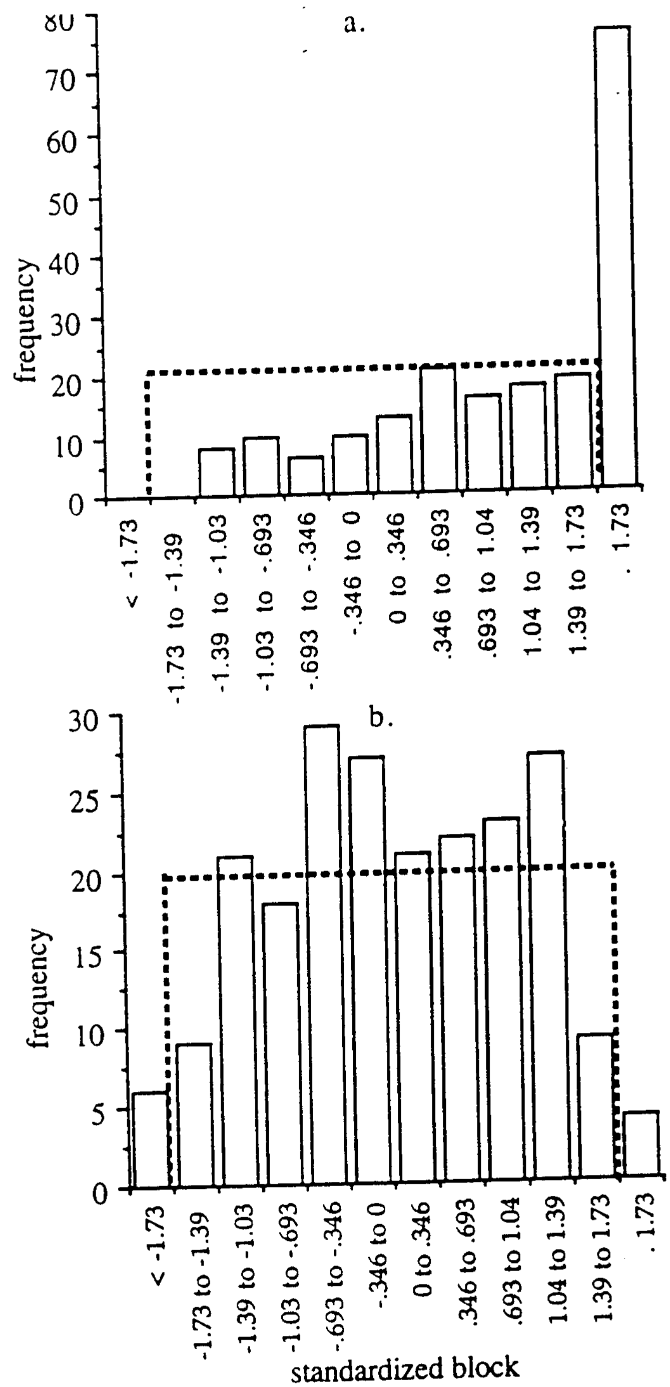


Figure 10. Frequency distributions of the parameter *a* of the logistics fits for each item. The values of *a* were standardized for each subject by computing z-scores based on the range from block 2 to the largest value of *a* for that subject (Panel a), and based on the observed means and standard deviations of *a* for that subject (Panel b). These standardized data were then collapsed across subjects to yield the distributions shown in the figure. Dashed rectangles in both panels represent the expected frequency distribution if the data follow a uniform distribution incorporating the parametric assumptions of each panel as discussed in the text (Experiment 1).

a variety of noise factors would be expected to cause the strengthening increment to be a little larger or a little smaller than average on a given learning trial. Under these conditions the central limit theorem requires that the number of trials necessary to reach a retrieval strength threshold (i.e., a transition midpoint) will be approximately normally distributed across items within each subject. To test both of these later possibilities, z-scores were computed based on the observed mean and SD of each subject's data, and the data were collapsed across subjects as shown in Figure 10b. The expected uniform distribution fit is again shown by the dashed lines for reference. Both distributions provide much improved fits. However, they can both also be rejected by a chi-square test; for the uniform distribution, $X^2(7, 216) = 20.6$, $p < .05$, and for the normal distribution, $X^2(7,216) = 20.58$, $p < .05$.[3]

Although the analyses above failed to identify an appropriate distributional model, they do provide for some important insights into the nature of the strategy transition process at the subject level. Neither a simple sequential learning model nor a more automatic and parallel (across items) strength accrual model appears to be appropriate. One possible alternative is that a single strengthening process is operating for each item, and that other more idiosyncratic learning processes are also operating for a subset of items. For example, some items may lend themselves more naturally than others to the use of mediational elaboration techniques (Ericsson, 1985). The available data from this experiment do not allow for any more specific conclusions regarding this or other possible accounts.

The CPL model predicts that the proportion of retrieval responses in the group data should be well fit by the average, collapsed across items and subjects, of the

[3]This and all subsequent chi-square tests for the normal distribution were performed on frequency counts derived by dividing the data into the following 10 categories. Eight of these categories were derived by dividing the data into equal intervals from a minimum of -2.0 to +2.0, and the remaining two categories covered z-scores below -2.0 and above + 2.0, respectively.

retrieval probabilities predicted by the logistic fits to each item. This predicted retrieval function is overlaid on the observed data in Figure 11. The fit is very good both

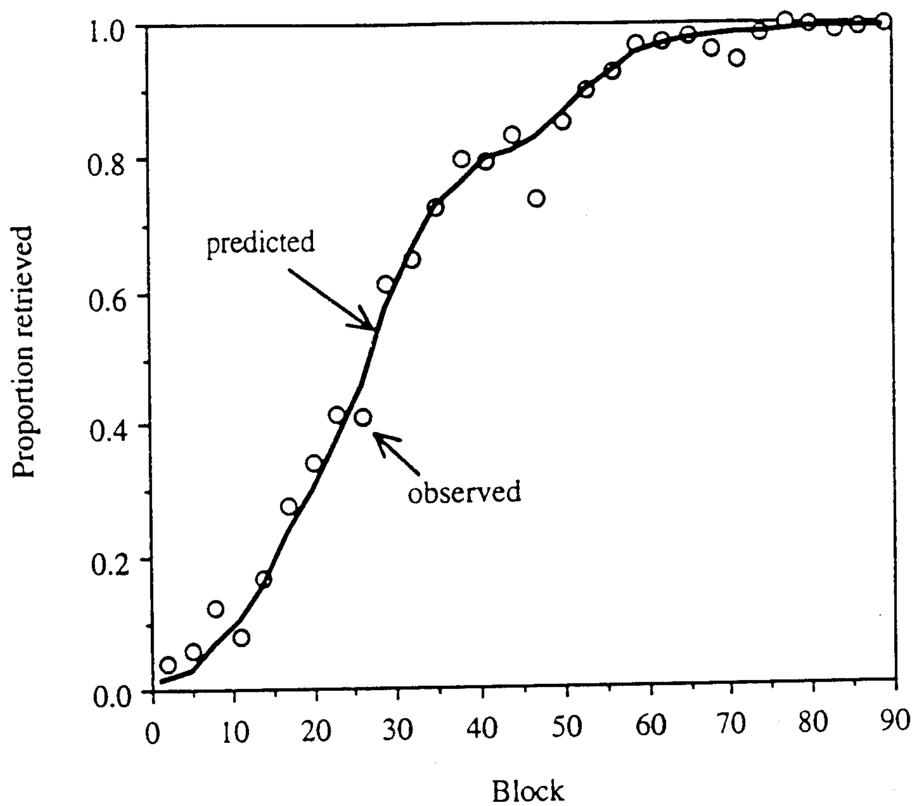


Figure 11. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *retrieval* strategy was reported as a function of block. Fitted line is the prediction of the CPL model (Experiment 1).

visually and statistically ($r^2 = .99$). The high accuracy of this fit is not surprizing given that the logistic function provides a very good model of the transition to retrieval at the item level. Nevertheless, this fit is useful because it confirms the ability of the logistic function to account for strategy transition data at the group level. In addition, as will be discussed later, this fit is important in constructing CPL fits to the overall RT and SD data.

Practice: CPL Fits to the RT and SD Data

One approach to evaluating RTs and SDs separately by strategy is simply to examine only the data on which strategy probes were collected. This approach, however, eliminates two-thirds of the data. An alternative approach which nearly triples the number of observations in each strategy category is to use the logistic fits to each item as a filter for selecting trials which with a high probability reflect algorithm or retrieval strategies. The following filtering procedure was employed for selecting algorithm trials. First, the practice blocks corresponding to predicted retrieval probabilities of .01 (*BLmin*) and .99 (*BLmax*) were computed based on the logistic fits to each item. All trials which occurred before *BLmin* for a given item were then selected as algorithm trials, with the exception of those trials on which the retrieval strategy was explicitly indicated by the strategy probing data. For block values between *BLmin* and *BLmax*, only trials on which strategy probing directly showed that the algorithm was used were selected. Although occasionally subjects reported using the algorithm on blocks above *BLmax*, theoretically there is a high probability that these cases reflects errors in the strategy response rather than actual use of the algorithm. Also, as discussed previously, the CPL model does not strictly predict that the power law will hold when outlier trials are included in the data (i.e., when transitions do not approximate step functions). For these reasons, trials on which subject reported using the algorithm on blocks above *BLmax* were excluded. The filter for retrieval trials was exactly analogous to that for algorithm trials, but in the reverse direction. The relatively small number of *other* responses were grouped into the *algorithm* strategy for this analysis.

Figure 12a shows the results for RTs, and Figure 12b shows the results for SDs, plotted in log-log coordinates. SDs for both the algorithm and retrieval strategies were plotted collapsed across consecutive three-block sequences of practice to reduce

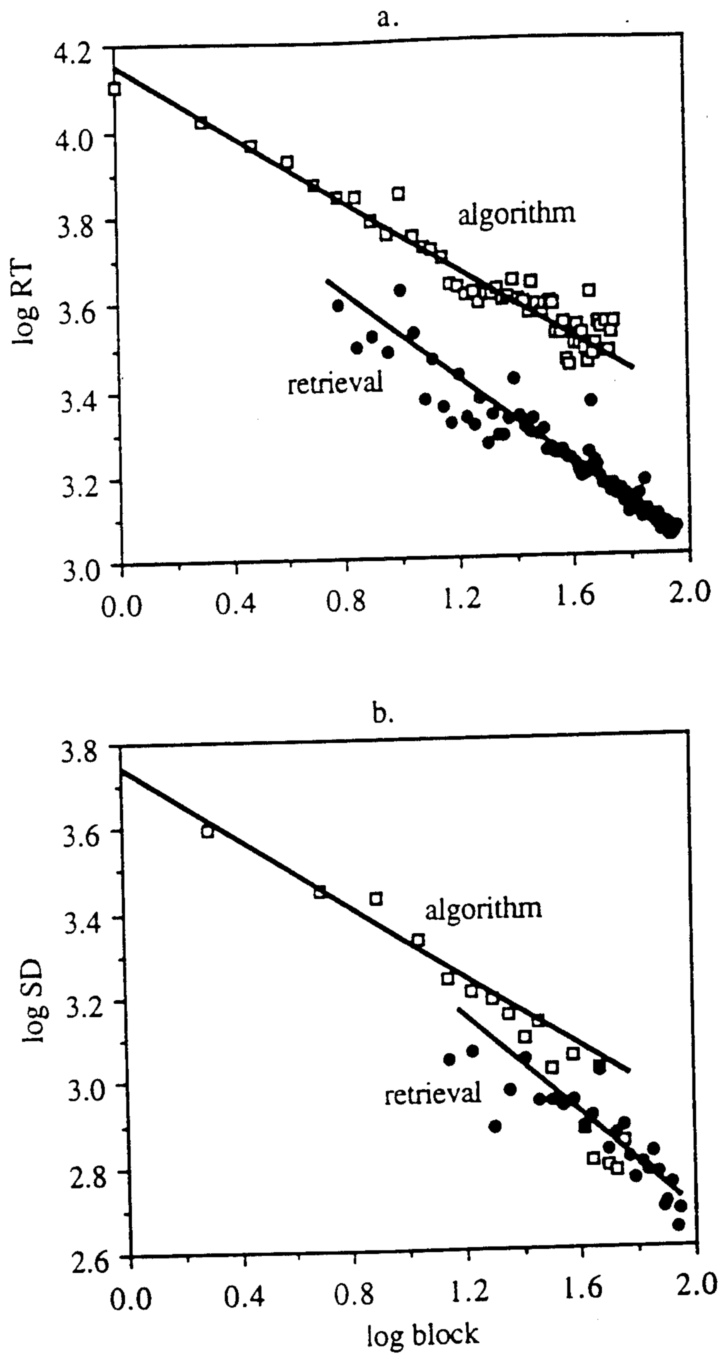


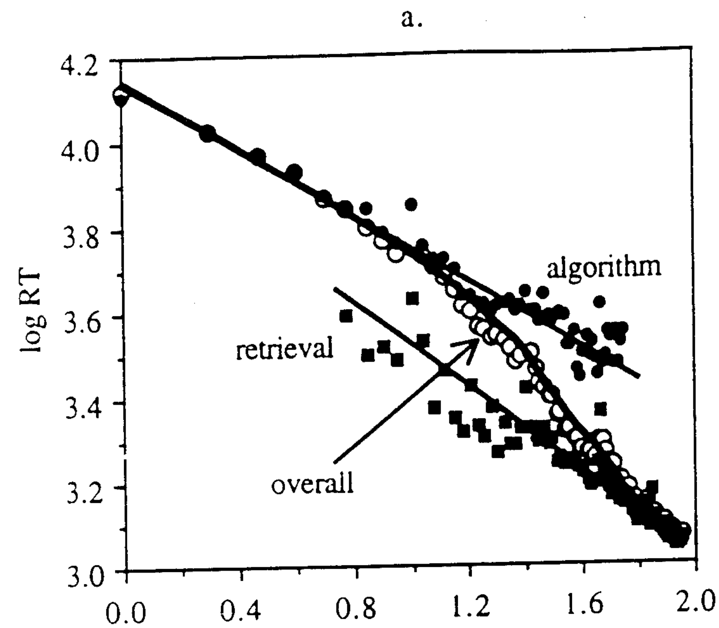
Figure 12. Log means (Panel a) and standard deviations (Panel b) of reactions times (collapsed across problems and subjects) plotted as a function of log block and strategy (algorithm or retrieval). Fitted lines represent best fitting power function for each strategy (Experiment 1).

noise in the figure. (Regression fits, however, were computed from the data at the individual block level.) Best fitting power functions are also shown for both strategies in both figures. These power function fits were limited to practice blocks on which all subjects contributed to the data, although the actual curve plotted extends the entire range of the data. There were no systematic deviations of the data from the fits for either the RT or SD for either strategy, with the exception of deviations for the SD on the last few algorithm observations and on the first few retrieval observations. As discussed earlier, these concave downward deviations are expected mathematically given the data collapsing approach used to construct the plots for the SDs.

CPL fits to the overall RT and SD were constructed by taking the antilog of the predicted values from each of the regression fits to the component strategies, squaring the SD values to yield variances, and then plugging these values, along with the predicted values of p as shown in Figure 11, into Equations 1 and 2. The predicted overall RT and SD (converted back to log-log coordinates) are shown in Figure 13 (a and b), overlaid on the observed overall RTs and SDs. Also shown for reference are the results for the component strategies. Generally the fits were quite good. The values of r^2 were very high (.98 for the RTs and .95 for the SDs) and more importantly there were only minor systematic discrepancies between the predicted and observed values. Clearly, these fits represent an improvement over those of the instance theory.

Practice: RT Results for the Nontransition Subject

A supplemental analysis was performed comparing the overall RT results for the 18 subjects who reported a transition to retrieval with those of the single additional subject who reported using the algorithm almost exclusively throughout the five practice sessions (see Figure 14). As predicted by the CPL model, the deviations from log-log linearity which are clear for the transition subjects are not evident at all for the



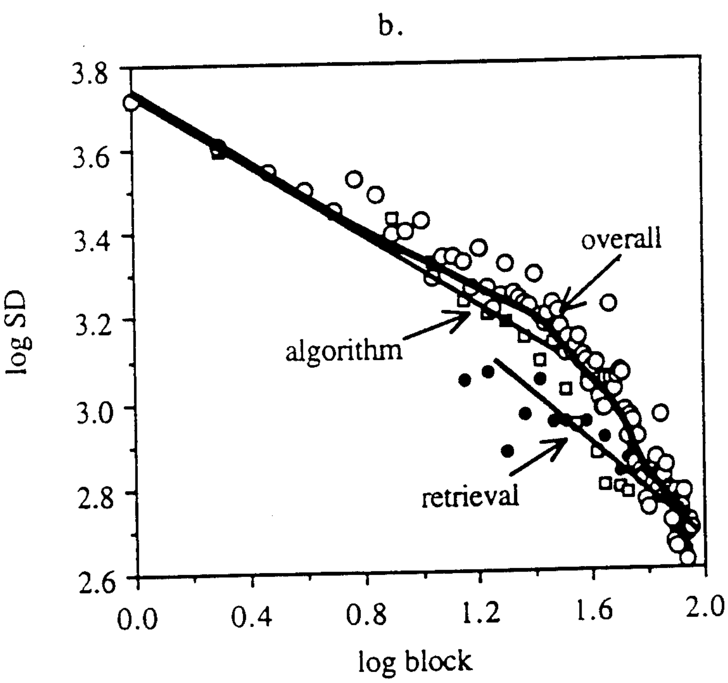


Figure 13. Log means (Panel a) and standard deviations (Panel b) of reactions times (collapsed across problems and subjects) plotted as a function of log block for both he overall data, and separately for the two strategies (algorithm or retrieval). Thin lines represent best fitting power function for each strategy, and thick lines represent CPL fits to the overall data (Experiment 1).

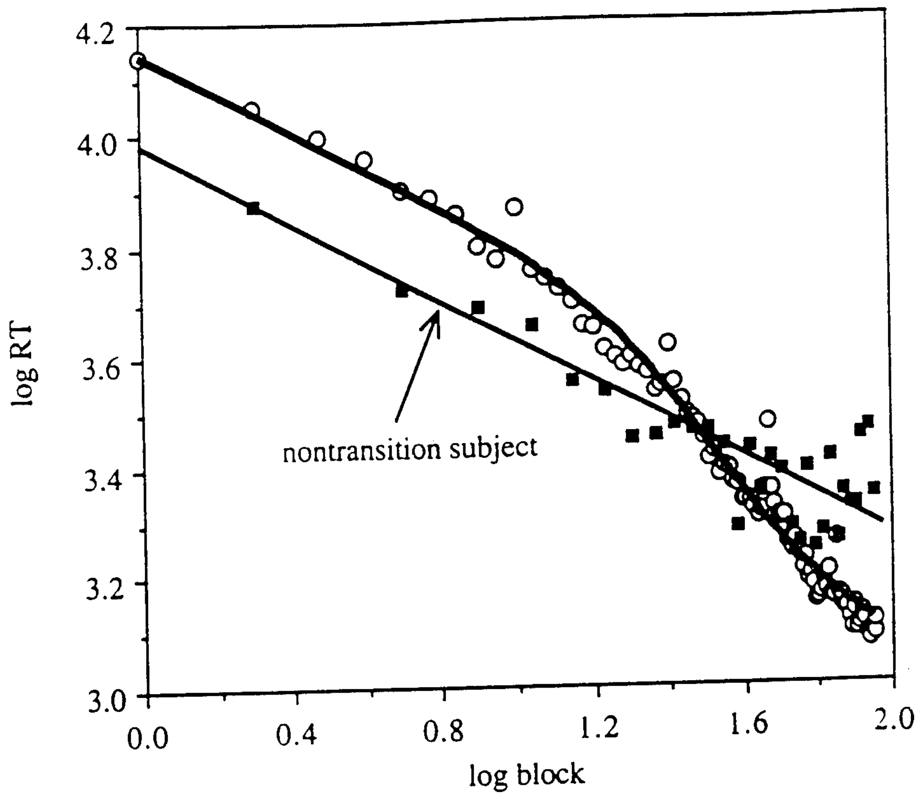


Figure 14. Log means of reaction times for the eighteen transition subjects, and for the nontransition subject, plotted as a function of log block. Fitted lines represent best fits of the CPL model (Experiment 1).

nontransition subject (because no strategy transition occurred for this subject, no deviations from the power function are predicted). Note also that although the nontransition subject was one of the fastest at solving problems initially, his performance at the end of practice was slower than that of each of the other 18 subjects. In addition to providing a between-subjects comfirmation of the predictions of the CPL model, these result provide strong evidence in support of the strategy probing data which indicate that the algorithm was used by this subject throughout practice.

Immediate and Delayed Tests

The probability of using each of the three strategies as indicated by the strategy probes for the three conditions on the immediate test is shown in Figure 15, collapsed across block. No-change problems show nearly total retrieval, not surprising given the complete transition to retrieval indicated for these problems during practice. In

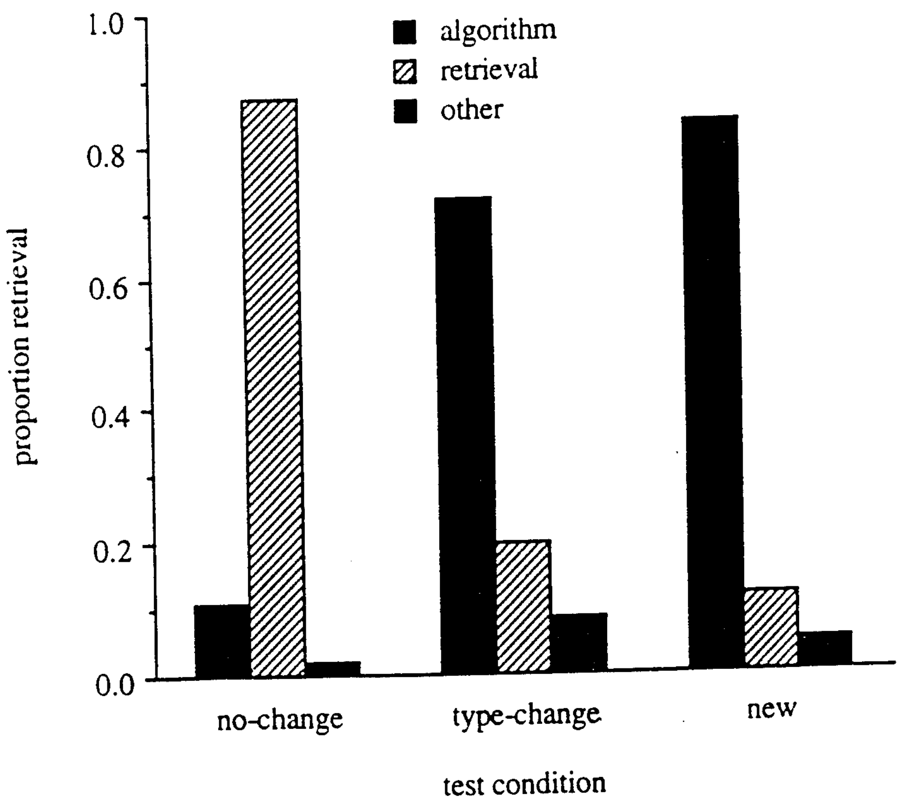


Figure 15. Proportion of trials (averaged over blocks of the immediate test and over all problems and subjects) on which subjects reported algorithm, *retrieval*, or *other* as a function of test condition (Experiment 1).

contrast, the algorithm was reported in most cases for new and type-change problems. A contrast performed on the proportion retrieved, comparing the no-change condition with the other conditions, was highly significant, $F(1,17) = 323$, $p < .001$, but a second contrast comparing the type-change and new problems conditions was not reliable, $F(1,17) = 3.08$, $p = .088$. Thus, the transition to retrieval was quite specific to the problems on which subjects practiced.

Error proportions and RTs at test showed similar results. The overall error proportions on the immediate test (collapsed across blocks) for the no-change, type-change, and new problems conditions were .024, .250, and .284, respectively. The difference between no-change problem on one hand, and type-change and new problems on the other hand, was strongly reliable, $F(1, 17)) = 53.8$, $p < .001$, but the difference between type-change and new problems was not reliable, $F(1, 17) < 1$. The RTs on the immediate test are shown in Figure 16 as a function of block and test

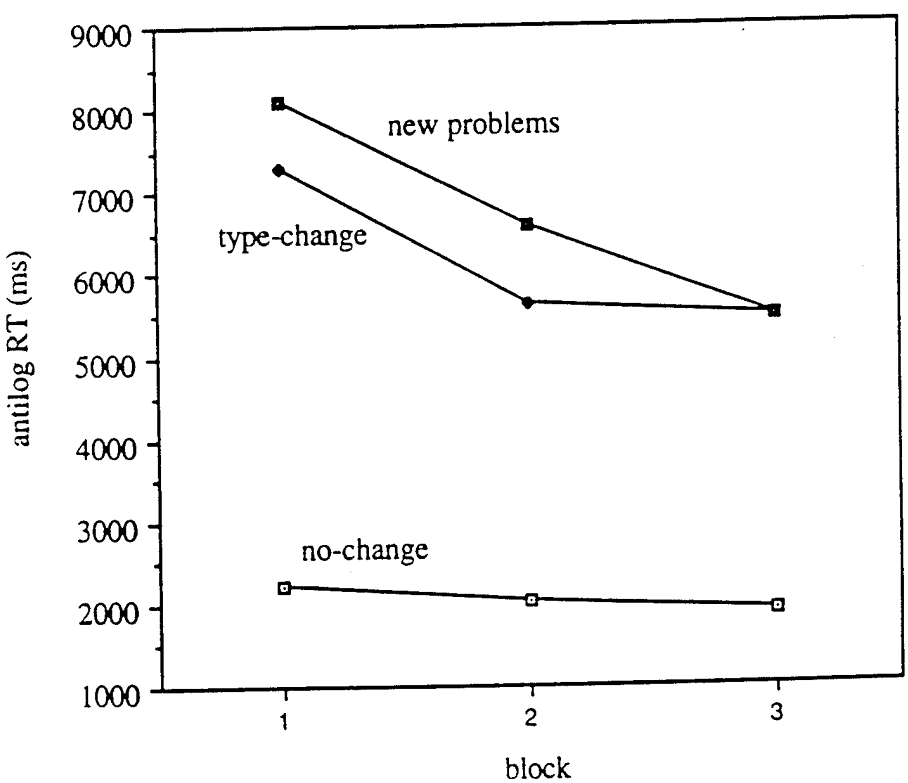


Figure 16. Antilog mean reaction time for correctly solved problems on the immediate test as a function of test block and test condition (Experiment 1).

condition. An ANOVA revealed reliable effects of test condition, $F(2, 17) = 42.6$, $p < .001$, block, $F(2, 17) = 6.68$, $p = .004$, and of the interaction between these variables, $F(4, 17) = 5.82$, $p < .001$. The interaction reflects greater speedup across the blocks of test in the new problems and type-change conditions than in the no-change condition. Contrasts showed that overall RTs in the no-change condition were reliably faster than in the type-change and new problems conditions, $F(1,17) = 84.05$, $p < .001$, but there was no evidence of any difference between the type-change and new problems conditions, $F(1, 17) = 1.21$, $p = .279$.

The algorithm RTs from practice, combined with the new problems RTs at test, provide a way to estimate the amount of speedup that reflects general speedup in the algorithm, and the amount that reflects speedup in executing the algorithm for specific problems. If all the speedup is general, then the RTs for algorithm trials on the last few blocks of practice on which they were reported should be roughly the same as the RTs for new problems at test. Alternatively, if the algorithm speedup is largely problem specific, then RTs for algorithm problems on the last few blocks of practice should be faster than the RTs for new problems at test. The RTs for algorithm trials during practice, on the last block on which the algorithm was reported at least 10% of the time, were around 3600 ms. Compare these to RTs for new problems on the first block of the immediate test of 8000 ms, and to algorithm RTs of around 13000 msec at the beginning of practice. These results suggest that some of the algorithm speedup is general, and some is specific to the problems on which subjects practiced. Neither of these forms of algorithm speedup are predicted by the instance theory, but they are both consistent with the CPL model.

An additional effect at test which should be noted is the slower RTs on no-change problems at test relative to the last block of practice. On the last block of practice, the antilog of the average log RTs for both problems types was around 1200 ms. At test, however, these RTs slowed to 2000 ms. Note that because the immediate test was given immediately after the last block of practice, this slowdown in RT cannot be attributed to forgetting. Analogous effects were observed by Rickard et al. (1994) in practice-transfer experiments exploring simple arithmetic skills (e.g., 4 x 7), although the effects in their experiments were much smaller (on the order of 100 ms). Rickard et al. (1994) interpreted this result in terms of the contextual interference concept introduced by (Battig, 1978). Following Battig, they speculated that the presence of new, unpracticed problems constituted a new context which interfered with retrieval of the practiced arithmetic facts. A similar account is plausible here, although the reason for the much larger scale of the effect in the current experiment is unclear. One possibility is that during practice subjects learned to associate numbers which were unique to one problem with the answer to that problem. For example, for subjects practicing on Problem Set 1 (see Appendix 1), the number 17 was present only in the problem *3 # 17 =* __, and thus was by itself a sufficient cue for retrieval of the answer to that problem. At test, however, the number 17 occurred in 6 problems (once in each problem set). Thus, reliable retrieval at test would require more global processing the the number configuration unique to each problem. The fact that subjects could still retrieve the answer relatively quickly at test indicates that this more global problem-answer association did form during practice. The finding that subjects were slower on no-change problems at test than at practice, however, suggests that this form of problem-answer association was at least supplemented during practice by simpler associations between presented numbers which were unique to a single problem, and the corresponding answer to that problem. Additional research is warranted to explore these hypothesized differences in associative structure in greater detail.

A comparison of RTs on the immediate and delayed tests, collapsed across blocks within each test, is shown in Figure 17. Solid lines represent overall results within each test condition. RTs for no-change problems on the delayed test were about half-way between RTs for no-change and new problems on the immediate test, indicating some skill retention. Nevertheless, the substantial increase in RT for no-change problems on the delayed test indicated a much greater loss in skill across the retention interval than we had observed in our previous work on arithmetic (see Fendrich et al., 1993, and Experiment 1 of Rickard, 1992). To investigate this finding

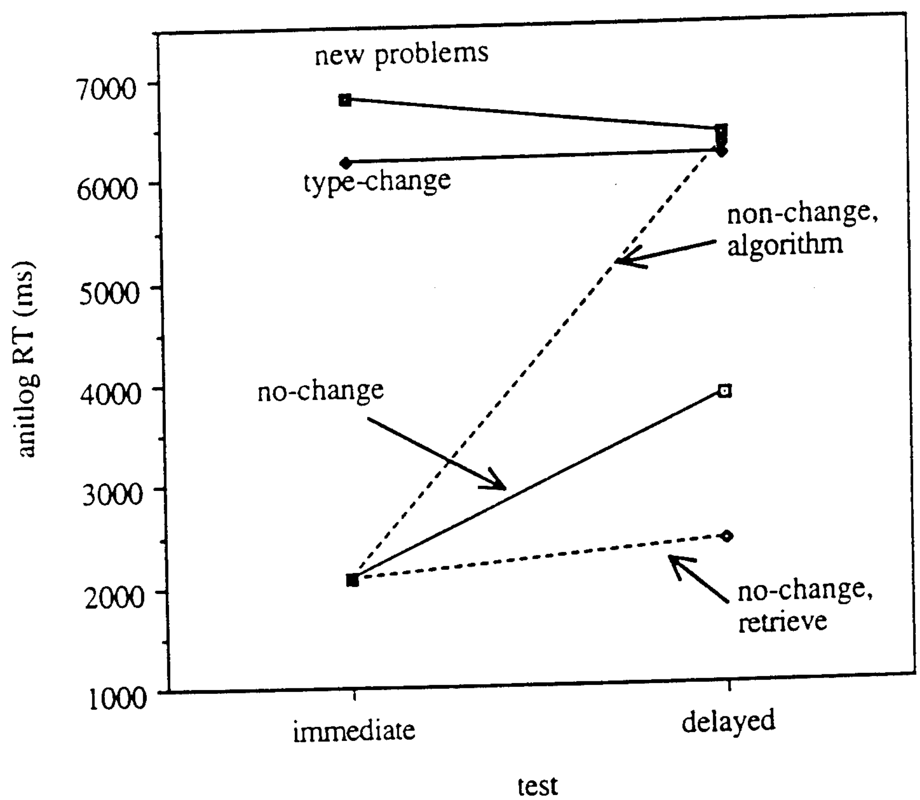


Figure 17. Antilog mean reaction time (averaged over block within each test) for correctly solved problems as a function of test and test condition. No-change problems on the delayed test are also plotted separately (as shown by the dashed lines) by strategy (Experiment 1).

further, we plotted the RTs for no-change problems on the delayed test separately by strategy, as shown by the dashed lines in Figure 17. When retrieval was the reported strategy for no-change problems on the delayed test, RTs were only slightly slower than for no-change problems on the immediate test, although this difference was reliable, $F(1, 17) = 9.43$, $p = .007$. When the algorithm was the reported strategy, the RTs were nearly exactly the same as those for new and type-change problems. Thus, training procedures which promote the use of an optimal strategy appear to contribute to the maintenance of training levels of performance on later tests of retention.

Summary

The following results of Experiment 1 provide strong support for the CPL model and also evidence against the instance theory: (a) the transition to retrieval was direct, with very few "other" responses, (b) the proportion of *retrieval* responses as a function of practice was well fit by the logistic function both at the item and group levels, (c) speedup in mean RT and reduction in SD clearly followed power functions within both the algorithm and retrieval strategies, (d) overall speedup and reduction in SD deviated from power functions in almost exactly the way predicted by the CPL model, and (e) the RT and SD data of the one subject who did not report a transition to retrieval did not deviate from the power function. The following results were also inconsistent with the instance theory (but not with the CPL model): (a) the rate parameters of 3 parameter power function fits to the RT and SD data were reliably different, (b) the proportion of *algorithm* responses as a function of practice was not well approximated by a power function, (c) the RTs for new problems on the immediate test provided clear evidence of both general and specific speedup in algorithm execution time.

CHAPTER 3

EXPERIMENT 2

The purpose of this experiment was to explore the generality of the CPL model using the alphabet arithmetic task of Logan (1988). This task differs from pound arithmetic in two important ways which would intuitively make it a better candidate for exhibiting parallel rather than nonparallel strategy execution. First, the alphabet arithmetic algorithm entails "counting" through the alphabet, a skill which is already highly practiced and efficient for adult subjects. Only modest additional practice might be needed to allow this algorithm to be executed while also attending to and executing the retrieval strategy (if this type of divided attention is possible). Second, it is conceivable that execution of these two strategies in alphabet arithmetic involves completely or partially independent cognitive or neural modules. Retrieval of a single recently acquired fact (such as the answer to an alphabet arithmetic problem) might involve different systems than does sequential retrieval of highly practiced "chained associations" such as the alphabet. For example, retrieving the answer directly in alphabet arithmetic may involve access to some generic fact retrieval module, whereas recitation of the alphabet may take place in a more specialized auditory memory module. Both of these possibilities appear to make the alphabet arithmetic task a good candidate for exhibiting parallel strategy execution. Thus, if the CPL model holds unambiguously in this task, then it is reasonable to infer that it is appropriate for a variety of tasks which exhibit a transition from algorithm to retrieval. In contrast, if it does not hold, then factors determining boundary conditions for nonparallel and parallel strategy execution will be suggested.

To assure comparability of this experiment with that of Logan (1988; see also Compton & Logan, 1991), the task was constructed as a verification task. Problems were presented with candidate answers (e.g., F + 3 = I; True of False?). The addend sizes used in this experiment, 3, 5, and 7, overlap with and also extend the addend size range of 2 to 5 used by Logan (1988). This use of a large range of addend sizes should provide a strong test of the ability of the CPL model and of the instance theory to account for the transition from algorithm to retrieval over a relatively wide range of algorithm difficulty. In Logan's (1988) alphabet arithmetic experiment, addend size 2 problems showed negligible deviations from power function speedup and reduction in SD. This result is potentially consistent with the CPL model under the reasonable assumption that the component power functions for the two strategies for addend size 2 problems had very similar parameter values; that is, under the assumptions that throughout practice the algorithm RTs were only slightly slower than were the retrieval RTs. As addend size increased, however, deviations from log-log linearity for both the mean RT and the SD became more pronounced. For addend 5 problems, these deviations were obvious. According to the CPL model, these increasing deviations from log-log linearity with increasing addend size reflect a progressively increasing "distance" between the retrieval and algorithm component power functions. Inclusion of a wide range of addend sizes (3, 5, and 7) in the current experiment , combined with strategy probing allowing for separate plots of algorithm and retrieval RTs and SDs as in Experiment 1, should provide a strong test of this interpretation.

According to the CPL model algorithm and retrieval strategies are executed independently of one another. Thus, there is no necessary reason under the model that either the rate with which the transition to retrieval will take place or the functional characteristics of performance across practice (e.g., the RTs and SDs) for the retrieval strategy will depend on addend size. Indeed, the simplest prediction of the CPL model is that all of these variables will be invariant across addend size. This prediction, however, is not necessitated by the model because it is possible that subjects may adopt different learning strategies for the different addend sizes. For example, subjects may concentrate on rehearsing the answers to addend 7 problems because these are the problems which are most time consuming to solve by way of the algorithm. Thus, the characteristics of retrieval-based performance as a function of addend size does not provide the basis for a test of the CPL model. Nevertheless, this empirical issue is important more generally to understanding learning mechanisms reflected in the transition from algorithm to retrieval and thus will be one focus of the data analyses discussed below.

Method

Subjects, Apparatus, and Materials

Twenty-one subjects from an introductory psychology course participated in the experiment for credit. Subjects were tested on Zenith Data Systems personal computers, programmed with the Micro Experimental Language (MEL) software (Schneider, 1988). Twenty-four problems (12 true and 12 false) were constructed (see Appendix B). Eight problems with the addend 3, eight with the addend 5, and eight with the addend 7. Four problems within each addend size were true, and four were false.

Procedure

There were four experimental sessions, the first three on Monday, Wednesday, and Friday of one week, and the fourth on Monday of the following week. Each session lasted 30-45 min. Subjects were tested in groups of up to 4. At the beginning of the first session, the subjects were introduced to the alphabet arithmetic task by way of one true and one false problem worked on a blackboard by the experimenter (neither of these problems were in the stimulus set). Subjects then performed the task independently at their own computer. During the first session, subjects performed 15 blocks of problems using the computer, where each block was one exposure to each of the 12 problems in the subject's practice set. Problems were presented one at a time in the middle of the screen. Subjects entered "True" or "False" using specially marked adjacent keys on the numeric keypad. Subjects were instructed to use either the pointer finger of both hands (one for true and one for false) or the pointer and index finger of one hand, whichever was more comfortable. The "True" and "False" keys were counter-balanced across subjects. Subjects were instructed to work as fast as possible while being accurate. They were told that they could rest briefly between blocks of problems. The subject's answer for each problem was collected. Strategy probes ("algorithm", "retrieval", or "other") were collected on one-third of the trials as in Experiment 1. The second, third, and fourth sessions consisted of 21, 24, and 27 blocks of problems, respectively, presented on the computers as described previously.

There was no test to evaluate transfer effects as in Experiment 1. There were two reasons why the transfer test was omitted. First, the test of Experiment 1 was conducted primarily to provide a new test of the identical elements model of arithmetic fact retrieval (Rickard et al., 1994). Because a verification format is used in the current experiment, it was not possible to generate transfer conditions analogous to the type change conditions of Experiment 1, which were central to testing the identical elements model. Second, Logan and Klapp (1991) have previously conducted a transfer test to new problems using the alphabet arithmetic task. They found that practice does transfer partially to new problems at test. As Logan (1988) acknowledges, this finding is inconsistent with the current version of the instance theory, which assumes that algorithm finishing times do not change with practice. Note, however, that these findings are consistent with the CPL model, which explicitly allows for speedup in algorithm execution times.

Results and Discussion

True problems were solved slightly faster and slightly more accurately than were false problems. These effects, however, did not enter into any interactions with other variables, and thus data were collapsed across the true/false distinction in all of the following analyses. Error rates for addend 3 problems were .058, .042, .044, and .045 in sessions 1, 2, 3, and 4, respectively. For addend 5 problems these values were .083, .099, .077, and .064, and for addend 7 problems they were .090, .072, .072, and .070. A 4 (session) by 3 (addend size) within subjects ANOVA performed on the proportion of errors indicated a reliable increase in error rates with increasing of addend size, $F(2, 20) = 7.22$, $p = .002$. There was no reliable effects of session, $F(3, 20) = .48$, $p = .699$, and no interaction of these two variables, $F(6, 20) = 1.77$, $p = .111$. All analyses reported below were limited to correctly solved problems.

The strategy probing results are shown in Figure 18, collapsed over subjects, problems, and addend size. Practice appears to have been successful in creating a transition to retrieval. By about block 60, *retrieval* was the reported strategy on nearly all trials. As in Experiment 1, there were very few *other* responses, suggesting that there were no intermediate stages in which some third strategy was used. A within-subjects ANOVA performed on the overall proportion of retrieval responses with a single factor of addend size (means = .794, .791, and .799 for addend sizes of 3, 5, and 7, respectively) indicated than the rate of transition to retrieval was not influenced by addend size, $F(2,40) < 1$.

Instance Theory Fits

A one-parameter power function (see Results and Discussion section of Chapter 2) was fit to the proportion of trials on which the algorithm was the reported strategy (collapsed across addend size). As shown in Figure 19, the fit was poor, yielding an r^2 of .746, and exhibiting systematic visual deviations from the data. Figures 20, 21, and 22 show the overall log RTs (panel a in each figure) and log SDs (panel b in each figure) for the three addend sized plotted as a function of log block. Also shown in these figures are the best fitting power functions as predicted by the instance theory.

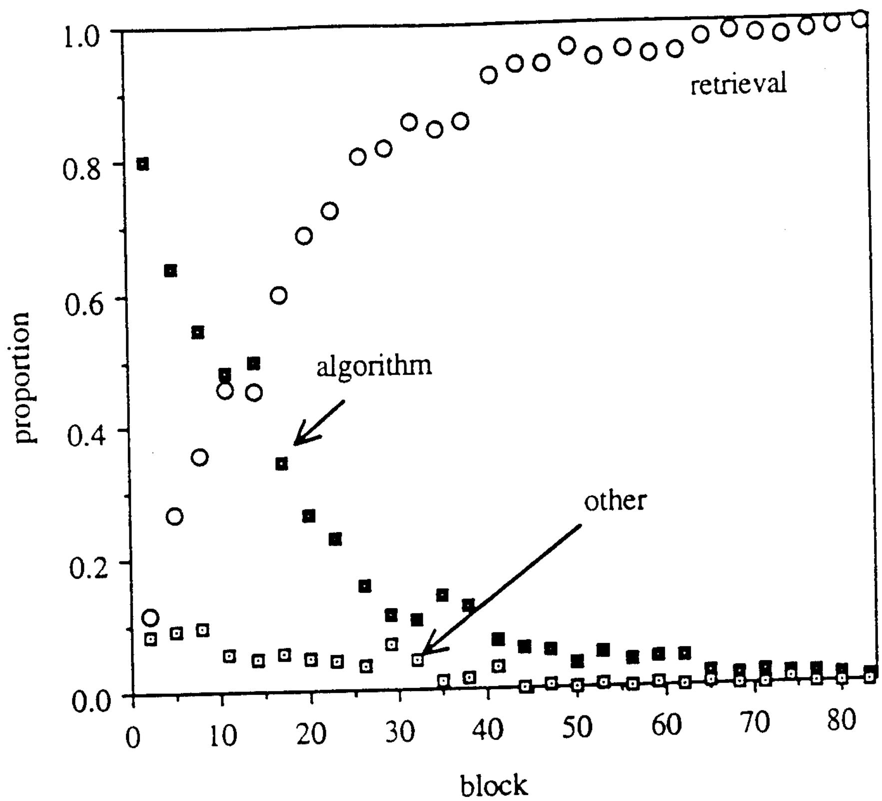


Figure 18. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *algorithm*, *retrieval*, and *other* strategies were reported as a function of block (Experiment 2).

Systematic deviations of the observed from the predicted values for both the RT and SD are clearly evident. Also, as was the case in the alphabet arithmetic data of Logan (1988), the deviations become larger with increasing addend size and are larger for the SD than for the RT.

The instance theory prediction of identical values for RT and SD power function rate parameters was evaluated separately for each addend size by computing the parameter estimates separately for each subject. For addend 3 problems, 15 of 21 subjects had larger rate estimates for SD than for RT. However, for addend 5 and 7 problems, 15 and 14 of the subjects, respectively, had larger rate estimates for the RT

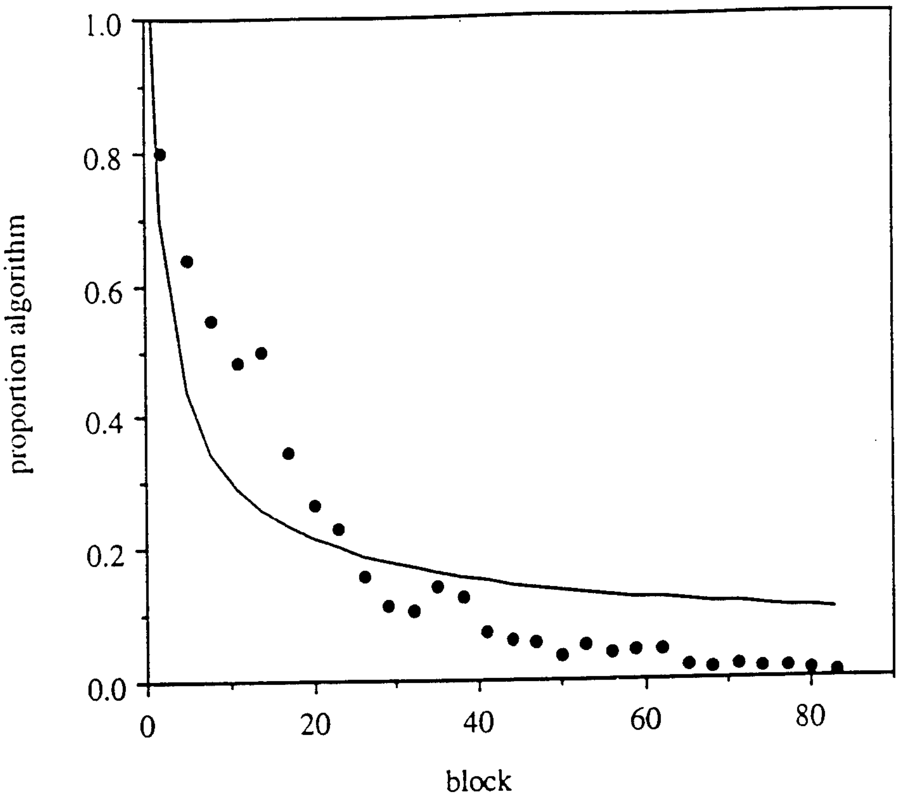


Figure 19. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *algorithm* strategy was reported as a function of block. Fitted line is a single-parameter power function as discussed in the text, which is a close approximation to the predictions of the instance theory (Experiment 2).

than for the SD. These effects for addend 3 and 5 problems were reliable by a binomial sign test (p's < .05). The apparent interaction between addend size and measure was confirmed by a within subjects ANOVA performed on the ranked rate estimates, $F(2, 40) = 13.2$, $p < .001$. Evidence of this same interaction is also present in the alphabet arithmetic results of Logan (1988, Experiment 4). For both true and false problems with addend sizes of 2, 3, and 4, the rate estimates for the SDs were larger than that for the RTs, but for both true and false addend 5 problems, rate estimates were larger for RTs than for SDs. Thus, evidence from two experiments now indicates that the rate

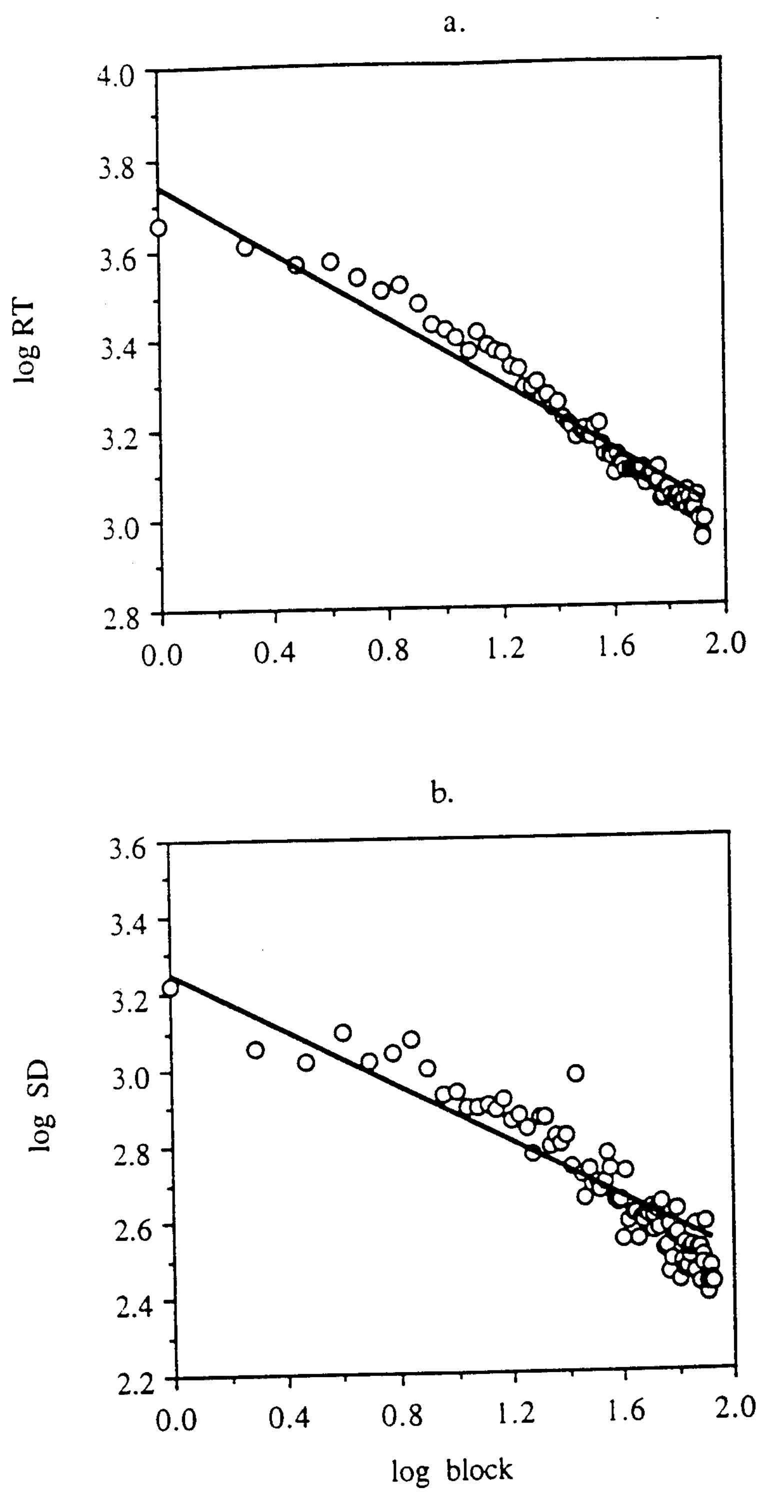


Figure 20. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 3 problems plotted as a function of log block. Fitted lines represent best fits of the instance theory (Experiment 2).

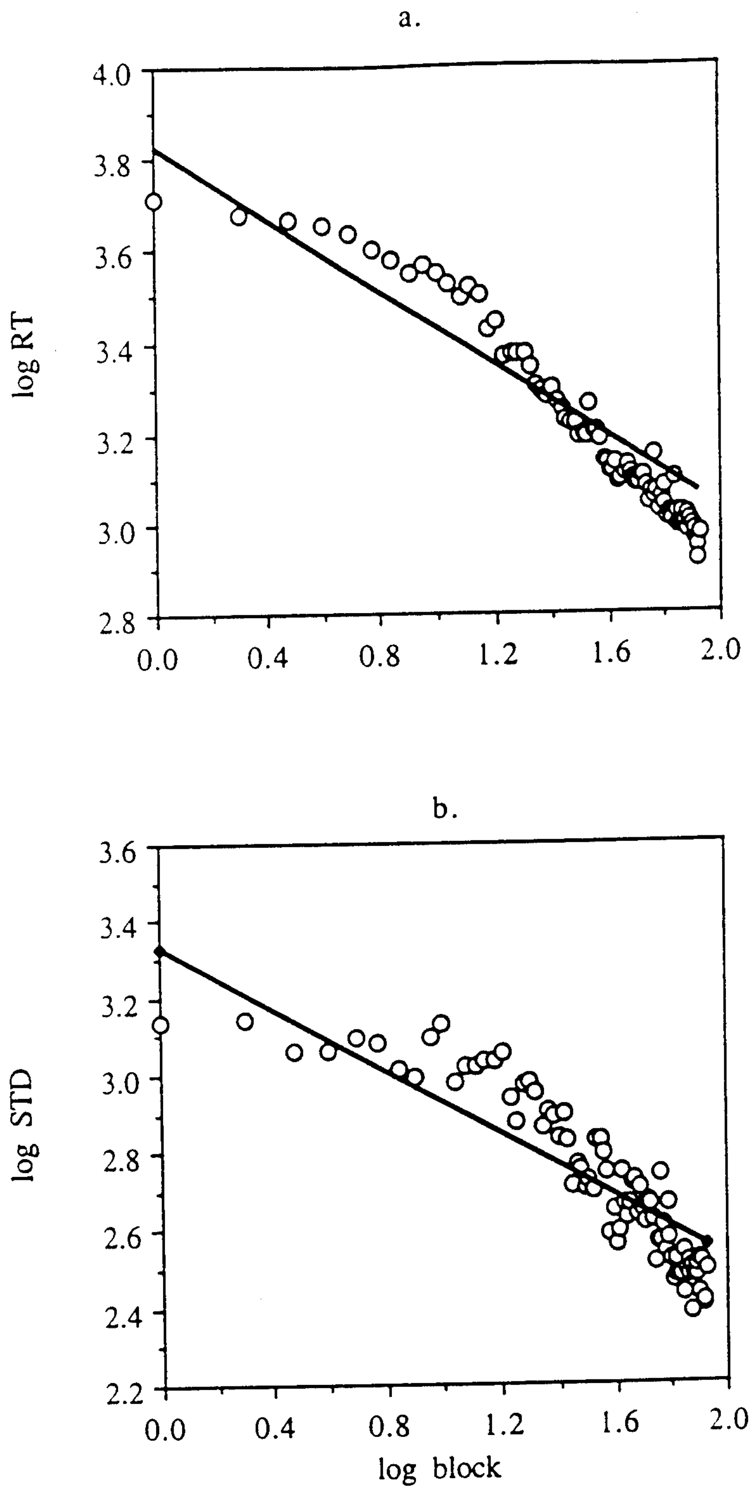


Figure 21. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 5 problems plotted as a function of log block. Fitted lines represent est fits of the instance theory (Experiment 2).

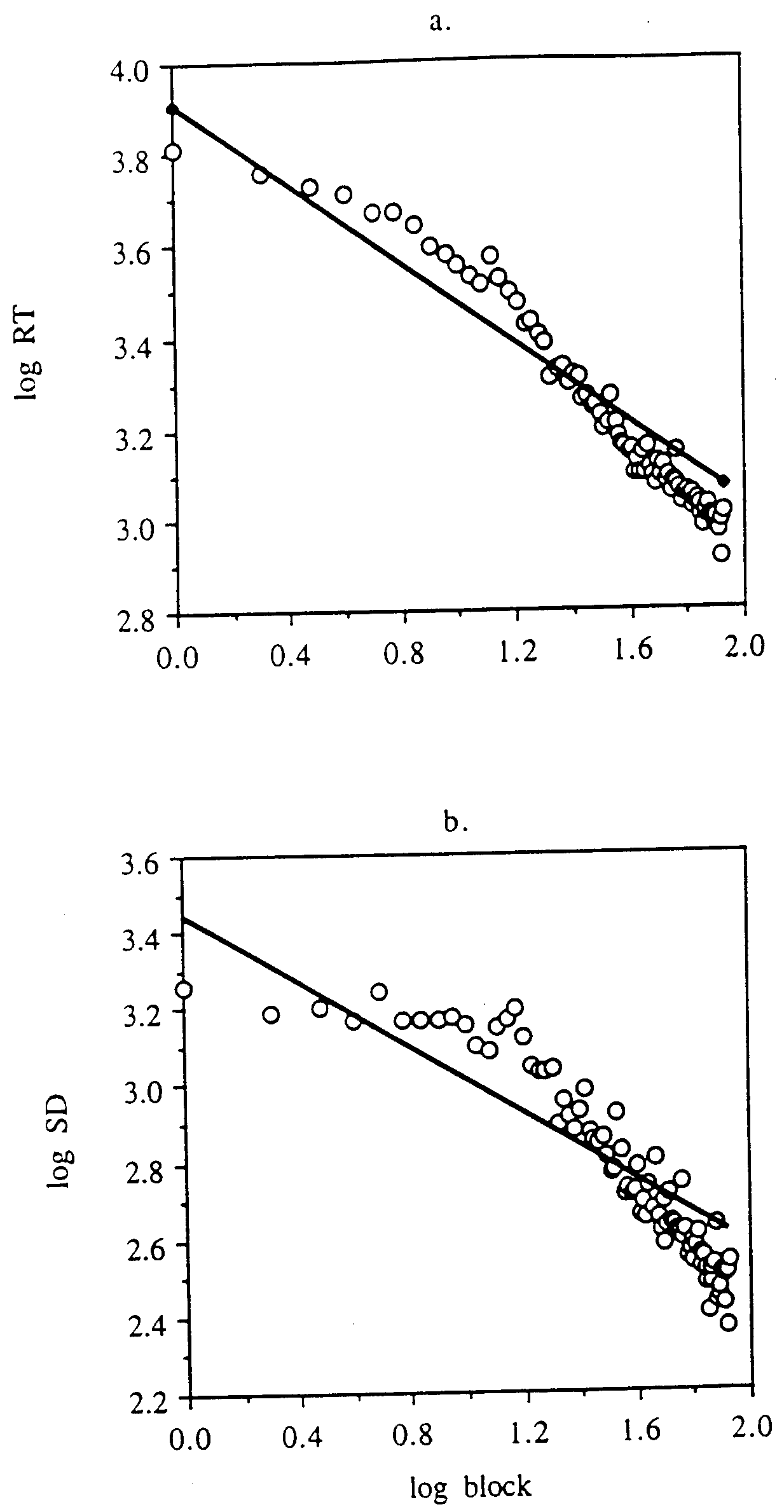


Figure 22. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 7 problems plotted as a function of the log of the practice block. Fitted lines represent best fits of the instance theory (Experiment 2).

estimates for the RTs increases faster than that for the SD as algorithm difficulty increases. This interaction contradicts the instance theory, which predicts that learning rates will be identical for the RT and SD regardless of algorithm difficulty.

CPL Fits: Strategy Transition Data

The CPL prediction that the probability of using the retrieval strategy at the item level conforms to a logistic function of practice was tested, as in Experiment 1, by fitting a logistic curve to each item separately for each subject. 247 of the 504 items (49%) showed a pure step function transition. The estimates for the retrieval midpoint parameter, *a*, for these problems were concentrated at early blocks of practice (mean = 14.3, SD =12.86). The remaining items did not exhibit step functions and evaluation of the quality of the logistic fits to these items was accomplished using the standardization approach discussed in Experiment 1. The results are shown in Figure 23. The logistic fit yielded an r^2 of .97, and exhibited no major systematic deviations from the observed data. Frequency distributions of the values of *a* and *spread* collapsed across items and subjects are shown in Figure 24 (a and b). The peak in the distribution of *a* occurred early during practice relative to Experiment 1, with a pronounced right skew. This result indicates that the transition to retrieval took place more quickly of average in this experiment than in Experiment 1, despite the fact that twice a many problems were seen during practice. One possible reason for this finding is that the alphabetic structure of alphabet arithmetic lends itself more easily to use of linguistic mediators than does the purely numeric structure of pound arithmetic (for a discussion role of mediators in skilled memory see Ericsson, 1985). Informal subjects interviews reported by Logan (1988) provide some evidence that mediators are indeed used in alphabet arithmetic. Additional research ,however, would be needed to verify this hypothesis. The distribution of *spread*, however, looked quite similar to that of

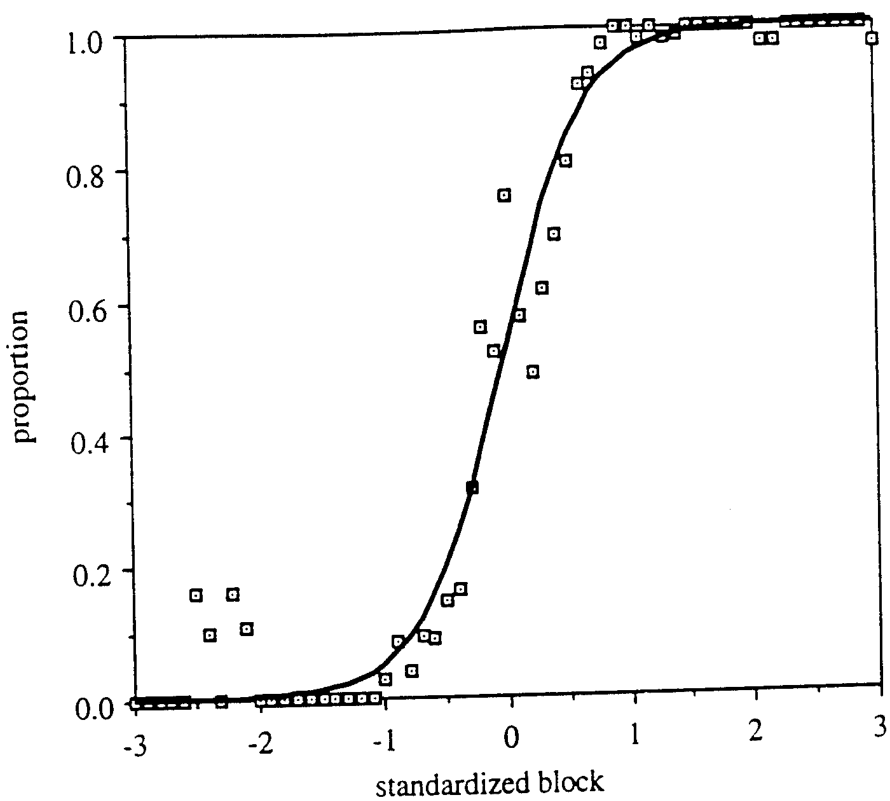


Figure 23. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *retrieval* strategy was reported as a function of standardized block. Problems for which the transition was a pure step function were excluded. Fitted line is the best fitting logistic function (Experiment 2).

Experiment 1, indicated that for most items the transition to retrieval occurred abruptly, approximating a step function.

The minimum, range, mean, and SD of *a* are shown separately for each subject in Table 2. Three candidate models of the retrieval midpoints were explored as in Experiment 1. The possibility that transition midpoints are uniformly distributed with a minimum value on the second block of practice and a maximum value corresponding to the last observed midpoint for each subject (see Figure 25 a) can be rejected easily, $X^2(7, 504) = 37.7$, $p < .01$. A second candidate model, which assumes that the distribution of retrieval midpoints is uniform for each subject, but with a minimum .

Table 2

Summary Statistics on Item Transition Midpoints in Experiment 2.

Subject	min	range	mean	SD
1	3.6	44.4	27.3	12.9
2	2	34.5	16.7	9.9
3	3.6	16.4	10.5	5.0
4	12.3	15.7	20.5	4
5	-.37	16.4	5.0	3.7
6	4.4	17.6	12.5	4.6
7	1.5	112.0	52.4	27.6
8	.2	27.7	13.9	6.9
9	2.6	147.3	29.4	33.3
10	6.4	25.1	18.5	6.3
11	2	12	8.3	3.9
12	1.5	16.5	10.4	5.5
13	13.9	34.0	30.9	8.4
14	-3.9	41.9	19.0	13.2
15	1.5	16.5	6.8	4.4
16	2.6	21.3	10.9	7.0
17	17.5	30.2	37.5	7.0
18	2.6	17.4	11.5	5.8
19	2.8	21.2	11.8	6.7
20	1.5	4.1	17.8	10.5
21	2.0	36.0	17.8	10.5

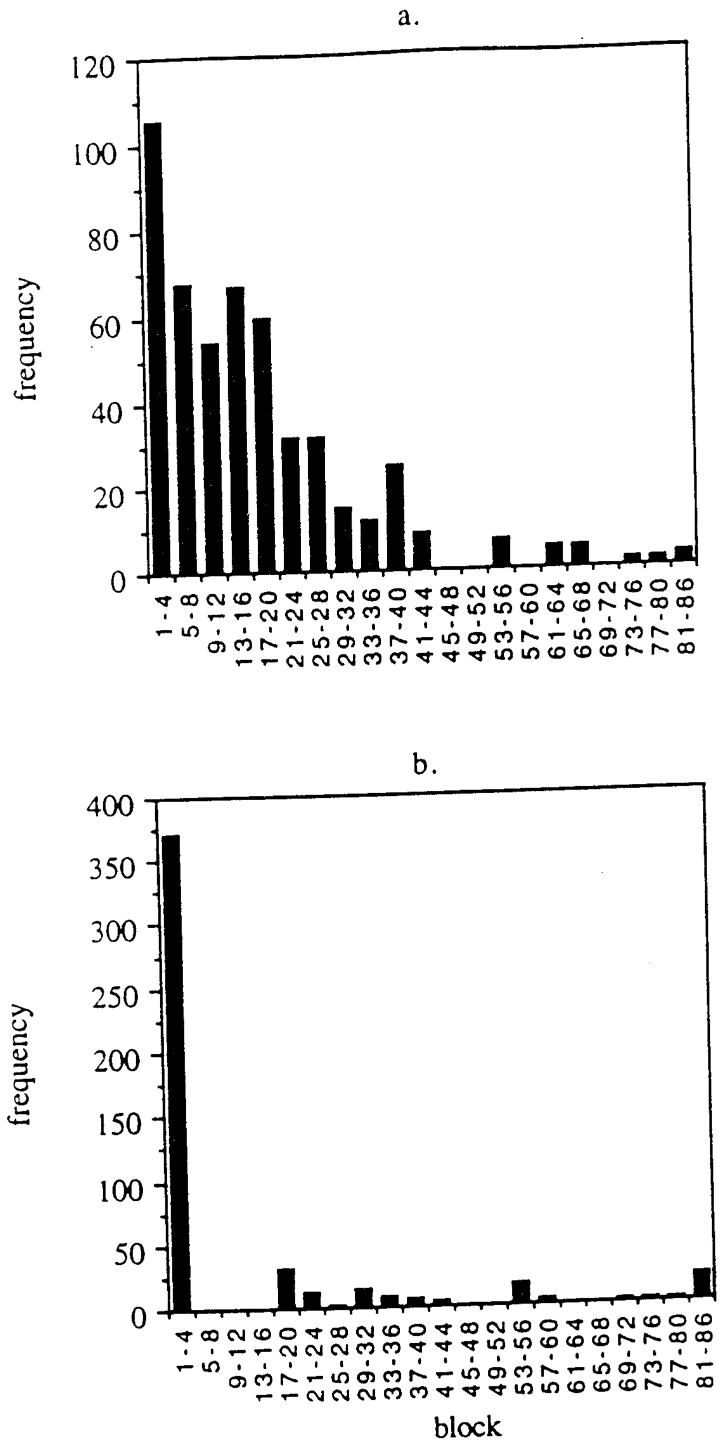


Figure 24. Frequency distributions for the parameter *a* (Panel a) and for the *spread* (Panel b) associated with the logisitic fits for all items (Experiment 2).

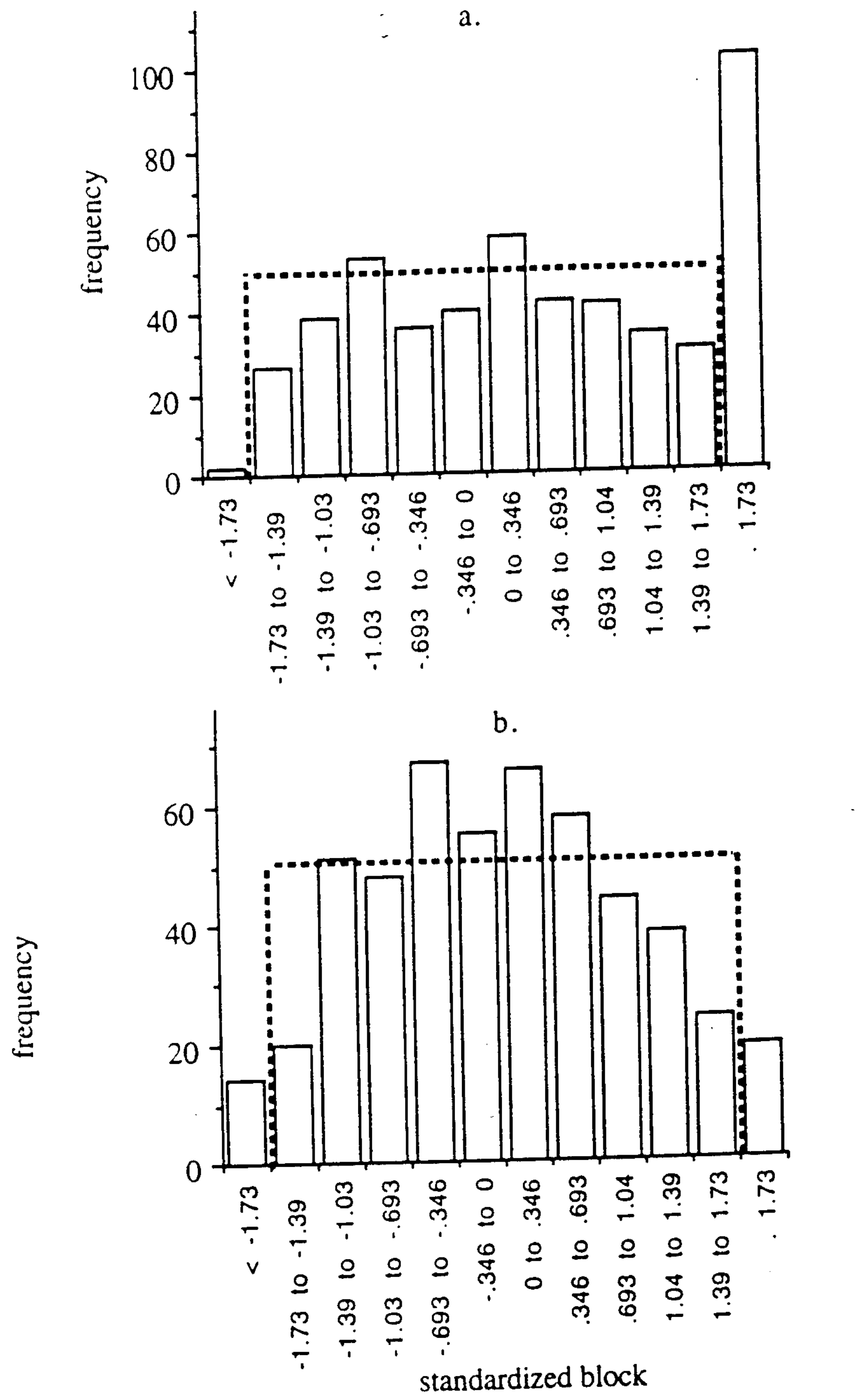


Figure 25. Frequency distributions of the parameter *a* of the logistics fits for each item. The values of *a* were standardized for each subject by computing z-scores based on the range from block 2 to the largest value of *a* for that subject (Panel a), and based on the observed means and standard deviations of *a* for that subject (Panel b). These standardized data were then collapsed across subjects to yield the distributions shown in the figure. Dashed rectangles in both panels represent the expected frequency distribution if the data follow a uniform distribution incorporating the parametric assumptions of each panel as discussed in the text (Experiment 2).

value which varies from subject to subject (Figure 25 b), can also be rejected, $X^2(7, 504) = 48, p < .05$, as can the normal distribution model, $X^2(7, 504) = 17.8, p < .05$ As in Experiment 1, a model somewhere "between" a normal and a uniform model appears to describe the distribution of transition midpoints at the subject level, suggesting the possibility that a collection of two or more distinct learning processes may be operating in this task domain.

The CPL fit to the overall proportion of retrieval responses, based on the logistic fits to the individual items, is shown in Figure 26. There were no systematic deviations from the predictions ($r^2 = .99$).

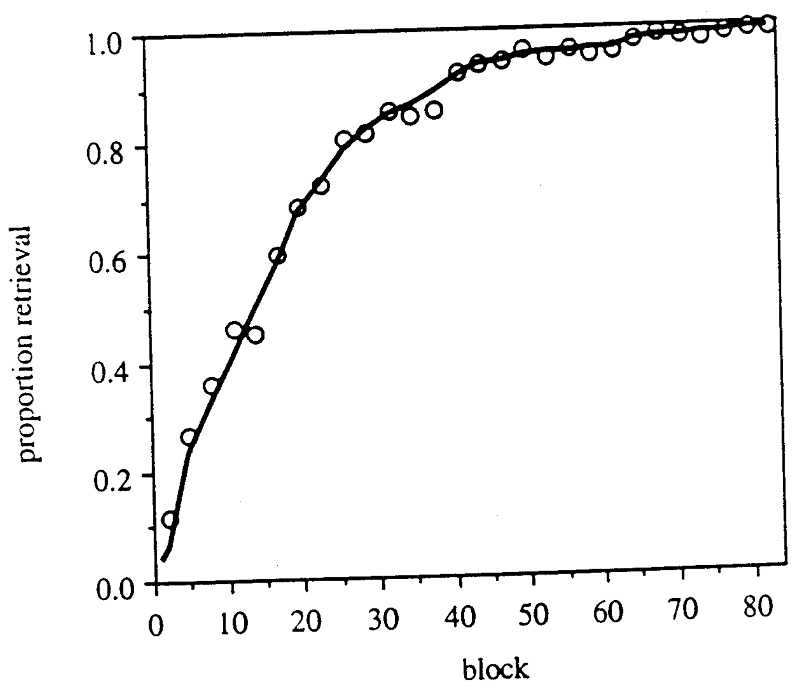


Figure 26. Proportion of strategy probing trials (averaged across consecutive three block sequences for all problems and subjects) on which the *retrieval* strategy was reported as a function of block. Fitted line is the prediction of the CPL model (Experiment 2).

CPL Fits: RT and SD Data

RTs and SDs corresponding to the algorithm and retrieval strategies were identified using the filtering approach discussed in Chapter 2. Figure 27 shows the algorithm RT results for all three addend sizes. As expected the RTs increases substantially with increasing addend size. Separate power functions were fit to data from each addend size. These and all subsequent component strategy fits were limited to blocks on which all subjects contributed an observation. The data in Figure 27

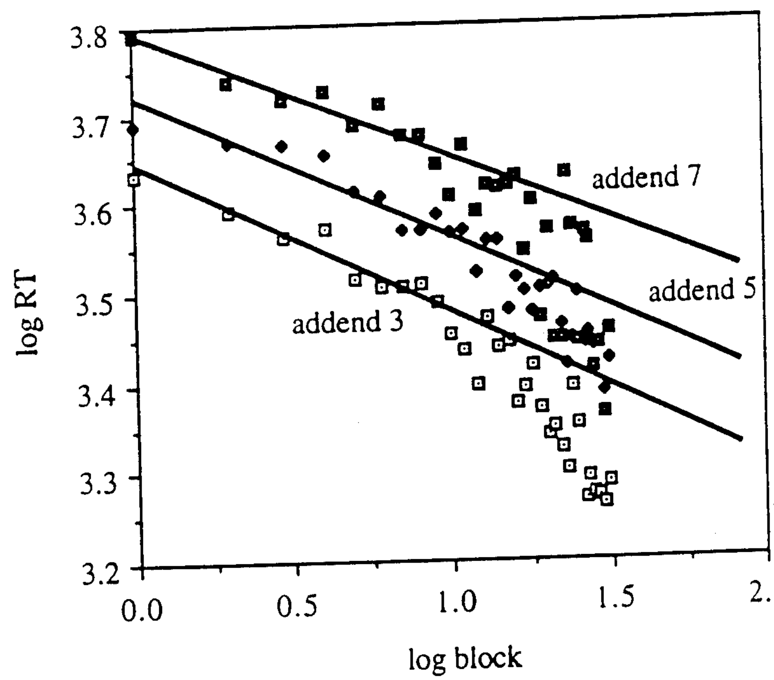


Figure 27. Log mean reaction time for the algorithm strategy plotted as a function of log block and addend size. Fitted lines are best fitting power functions based on data points to which all subjects made a contribution (Experiment 2).

conform closely to the power function fits for about the first 15 blocks of practice, but beyond that point they show clear deviations from the predictions for all three addend sizes. These results are potentially problematic for the CPL model and they will be discussed in more detail later.

An analysis of covariance (ANCOVA) with a continuous factor of log block and a categorical factor of addend (3, 5, or 7) was performed on the log RTs for the retrieval strategy to investigate whether addend size predicted retrieval-based RTs. There was a reliable effect of log block, $F(1, 20) = 489$, $p < .0001$, but there was no reliable effect of either addend, $F(2, 40) = .43$, $p = .66$, or the interaction log block by addend, $F(2, 40) = .63$, $p = .54$. Because addend did not predict retrieval performance, I collapsed across this variable and fit a power function to the overall retrieval data as shown in Figure 28. Clearly, power function speedup does hold for these data.

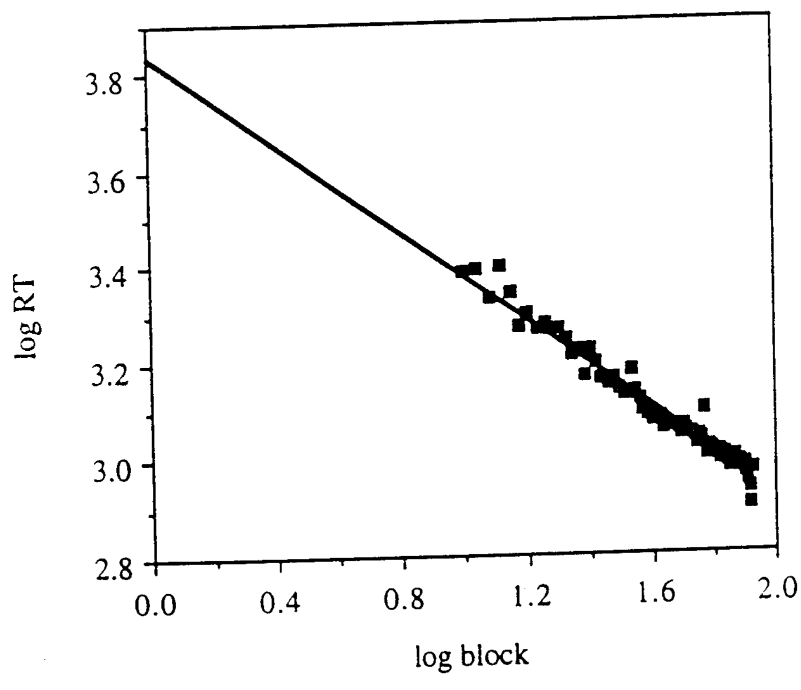


Figure 28. Log mean reaction time for the retrieval strategy plotted as a function of log block and collapsed over addend size. Fitted line is the best fitting power function based on data points to which all subjects made a contribution (Experiment 2). sizes. These deviations are potentially problematic for the CPL model. I will discuss them in more detail later.

Figure 29 (a, b, and c) shows the algorithm SD results separately by addend size. Power functions provide good fits to these data for about the first 30 blocks of practice, but beyond that point the data exhibit a concave downward deviation from linearity. This deviation from linearity is at least potentially accounted for, however, as bias in the SD estimates caused by the data collapsing methodology (see discussion in Chapter 1). Also, for the majority of observations, the power function does hold to a reasonable approximation.

An ANCOVA, identical to the one discussed above for retrieval RTs, was performed on the log SDs for the retrieval strategy to investigate whether addend size predicted retrieval-based SDs. There was again a very reliable effect of log block, $F(1, 20) = 69.5$, $p < .0001$, but there was no reliable effect of either addend, $F(2, 40) = .08$, $p = .92$, or the interaction log block by addend, $F(2, 40) = .06$, $p = .94$. Because addend did not predict retrieval SDs, I collapsed across this variable and fit a power function to the overall retrieval data as shown in Figure 30. With the exception of a few outliers on the lefthand side of the scatterplot, the power function provides a very good account of reduction in SD for these data.

Now consider once again the systematic deviations from linearity at all three addend sizes for algorithm RTs (Figure 27). The critical question for determining whether this effect is problematic for the CPL model is that of whether the effect reflects a correlation between transition midpoint and RT in the grouped data (see discussion in Chapter 1). Preliminary analyses revealed only mild trends toward a transition midpoint by RT correlation at the subject level. Thus, any correlation of this sort must, if it exists, occur at the item level within each subject. This possibility was investigated with the following analysis. First, power functions were fit to the log RT data for algorithm trials for each item for each subject, for a total of 504 fits. Then the log RT data for each item on each block of practice was standardized by subtracting from it the predicted intercept and the predicted value of the slope multiplied by the appropriate value of the log of the practice block. If the RT data at the item level

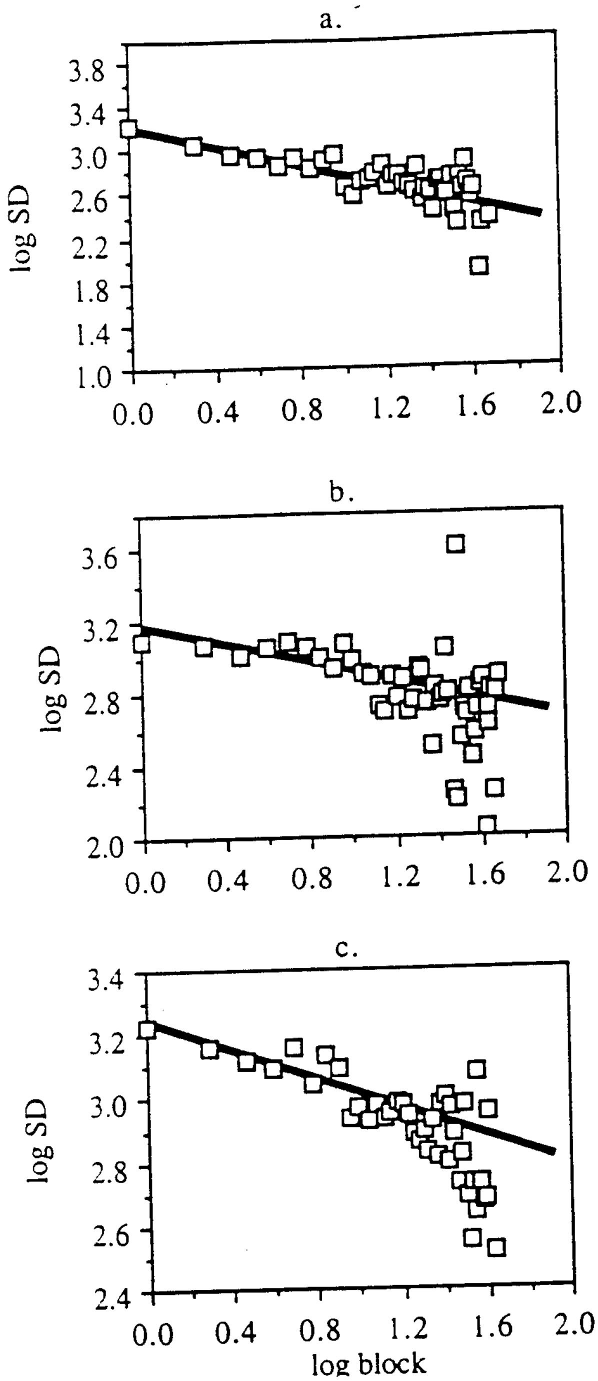


Figure 29. Log standard deviation of the reaction time for the algorithm strategy plotted as a function of log block. Addend sizes of 3, 5, and 7 are shown in Panels a, b, and c, respectively. Fitted lines are best fitting power functions based on data points to which all subjects made a contribution (Experiment 2).

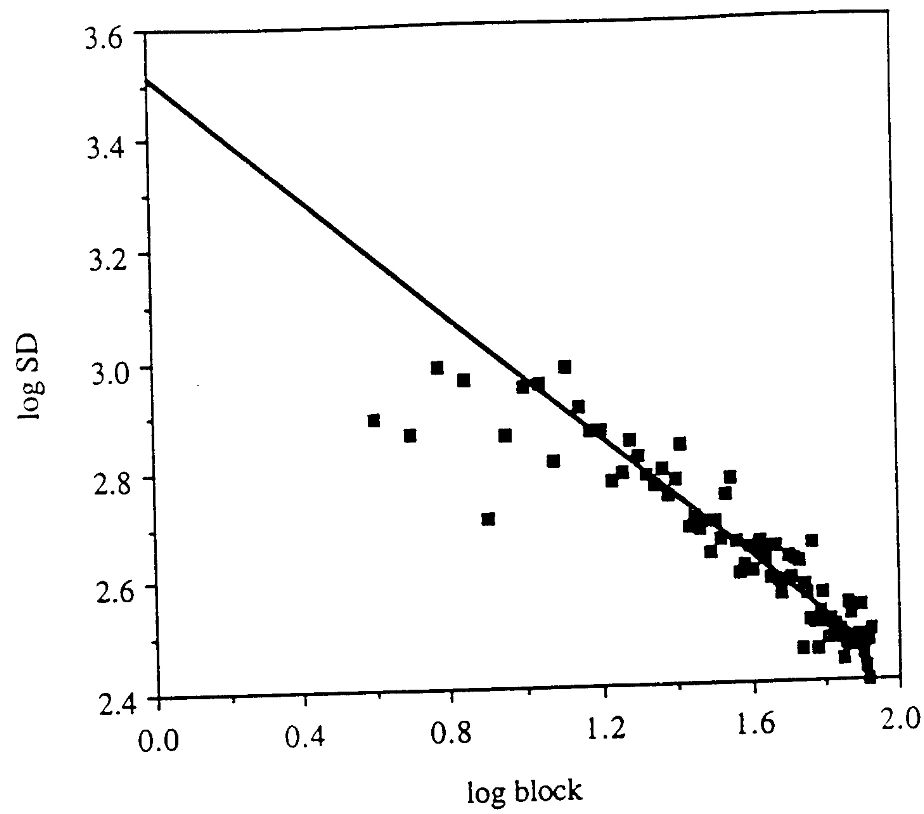


Figure 30. Log standard deviation for the retrieval strategy plotted as a function of log block and collapsed over addend size. Fitted line is the best fitting power function based on data points to which all subjects made a contribution (Experiment 2).

follows a power function, then each of these standardized RTs should, on average, have an intercept and a slope of zero. If the power function does not hold at the item level, then systematic deviations from these predicted intercept and slope values will be apparent. These standardized algorithm RT data, averaged across subjects and items, are shown in Figure 31. The data clearly have intercept and slope which are very close to zero. Indeed, least square second-order polynomial regression fit, shown in Figure 30, yielded an r^2 of only .0002.

The analysis discussed above indicates that the deviations from linearity evident in the algorithm RT data of Figure 27 reflect not a systematic deviation from the power function at the item level, but rather some form of transition midpoint by RT

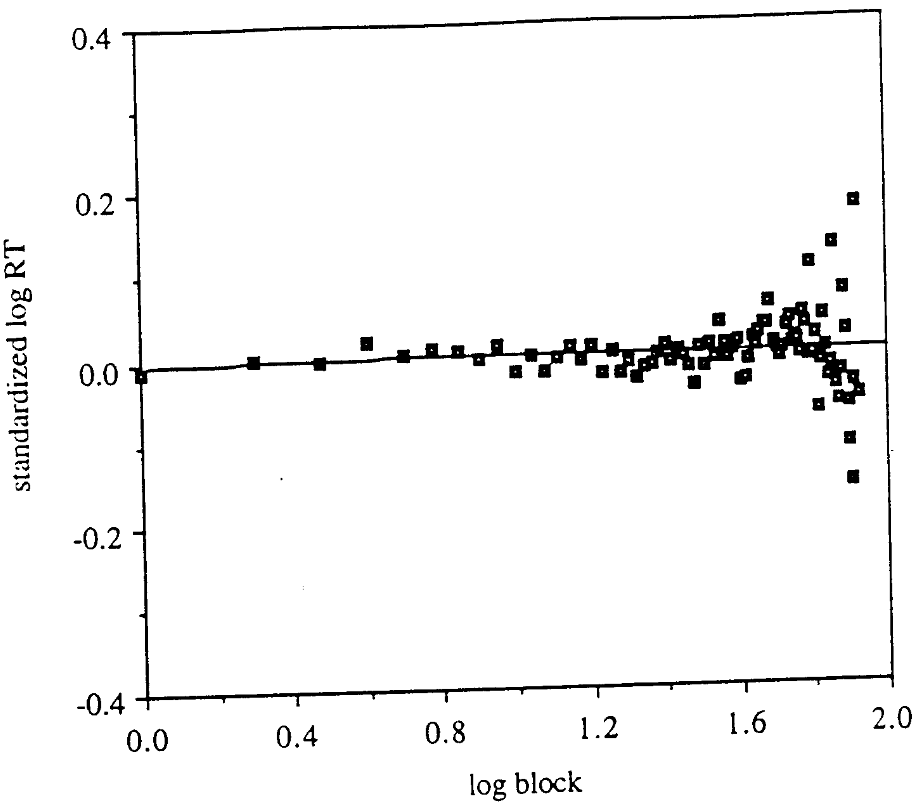


Figure 31. Standardized log mean reaction time for the algorithm strategy plotted as a function of log block. The fitted line represents the best fitting second-order least squares regression equation (Experiment 2).

correlation. Supplementary analyses indicated that a correlation does exist between the slope of the RT and the retrieval midpoint (the estimate of a in the logistic fits to each item); the shallower the slope (the slower the rate of speedup with practice), the faster the rate of transition to retrieval. This correlation can be corrected for mathematically to provide improved algorithm RT fits to the group data by applying the following equation:

$$RT_{avg} = \text{SUM [log RT (i)} * \text{p(i)/SUM(p)]}, \tag{4}$$

where RT_{avg} refers to the average (across all subjects and items) predicted algorithm log RT for a given block of practice, *log RT(i)* is the predicted algorithm log RT for item i based on the power function fits, $p(i)$ is the probability that the algorithm will be used for item i, based on the logistic fits to each item, and $SUM(p)$ is the sum of the probabilities that the algorithm strategy will be used across all items, also based on logistic fits to each item. Consider the simple case in which the algorithm is used with probability, $p(i)$, equal to 1 for all items. In this case, Equation 4 reduces to a simple average across items. In the case in which step function transitions occur for all items, then Equation 4 is simply an average of those items which are being solved using the algorithm at any given time. In the most general case in which $p(i)$ can vary between 0 and 1, Equation 4 is a weighted average in which the weighting factor is the algorithm probability for each item. Fits to the algorithm RT data based on Equation 4 are shown in Figure 32. The fits are clearly much improved for each addend size, although there

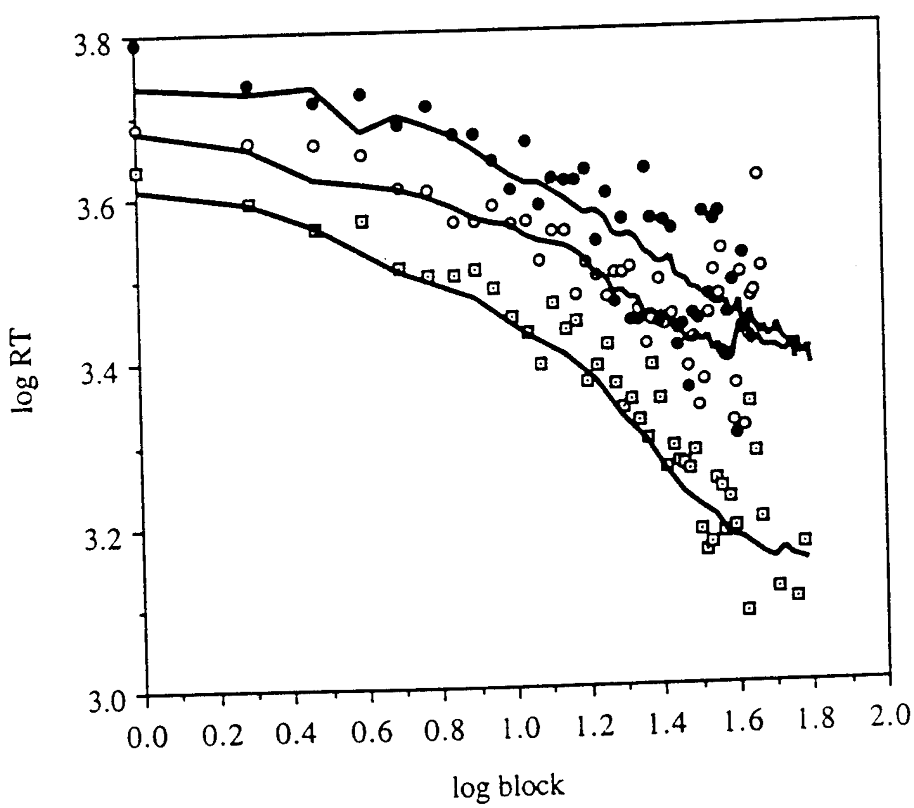


Figure 32.. Log mean reaction time for the algorithm strategy plotted as a function of log block and addend size (Experiment 2). Fitted lines are best fitting adjusted power functions based Equation 4.

is some systematic underestimation of the log RTs for addend size 3 problems.

Overall fits to the RT and SD were constructed as described in Experiment 1 (using the corrected fits for the algorithm RT discussed above). The predicted and observed overall RTs for each addend size, along with data and fits for the component strategies, are shown in Figures 33, 34, and 35. Panel a of each figure shows RTs and panel b shows SDs. Generally the fits were quite good overall (the r^2 values for addend sizes 3, 5, and 7 RT fits were .96, .94, and .94, respectively, and these values for the SD fits were .88, .82, and .84, respectively), although there was some relatively minor systematic deviations of the predicted from the observed values in some of the plots. Clearly, though, these fits represent a substantial improvement over those of the instance theory.

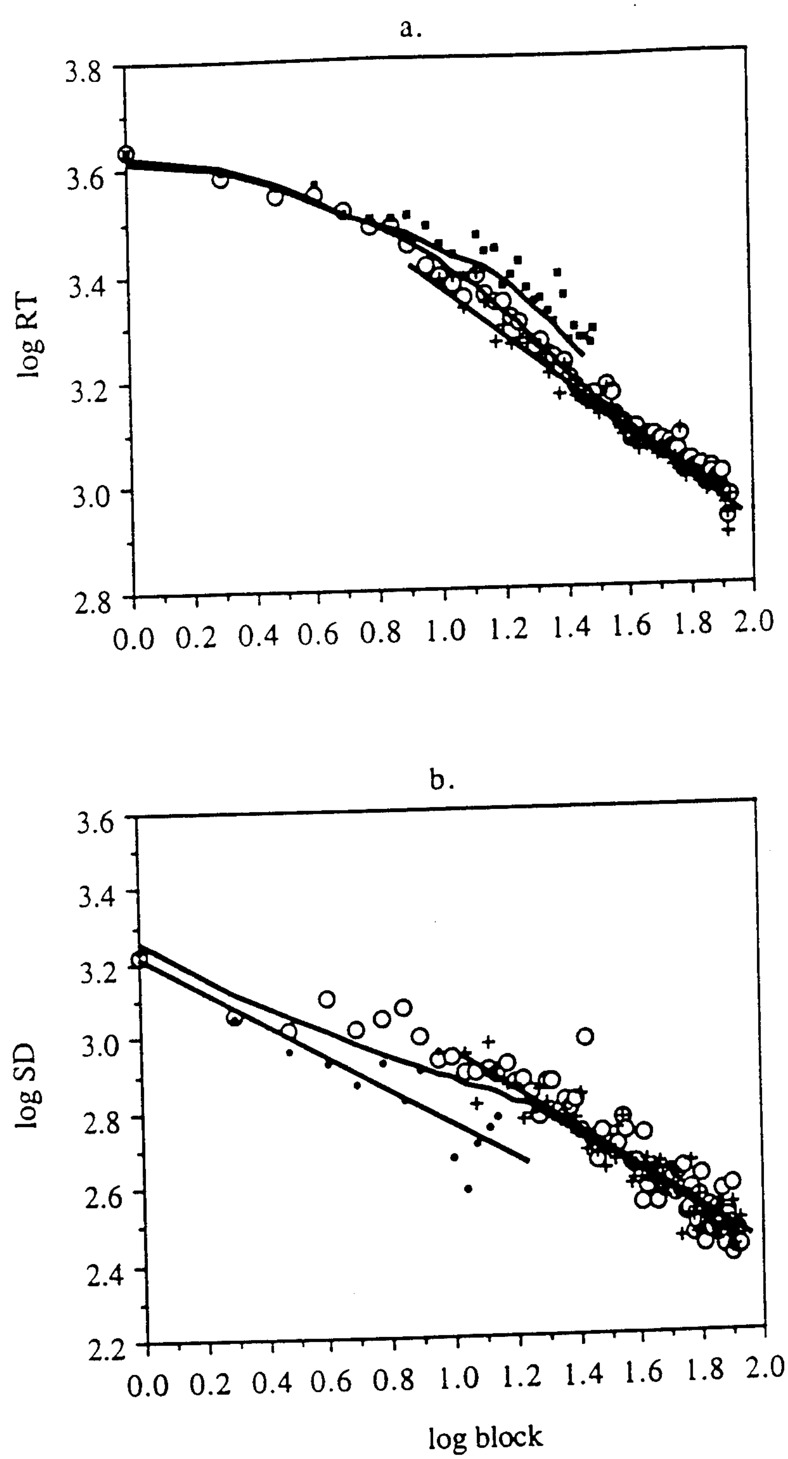
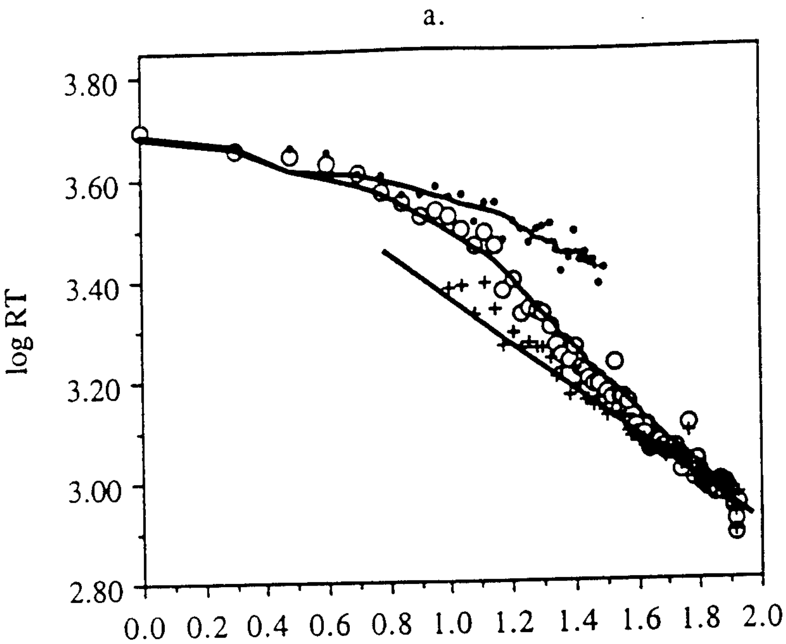


Figure 33. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 3 problems plotted as a function log block for both the overall data, and separately for the two strategies (algorithm or retrieval). Thin lines represent best fitting power function fits to each strategy, and thick lines represent CPL fits to the overall data (Experiment 2).



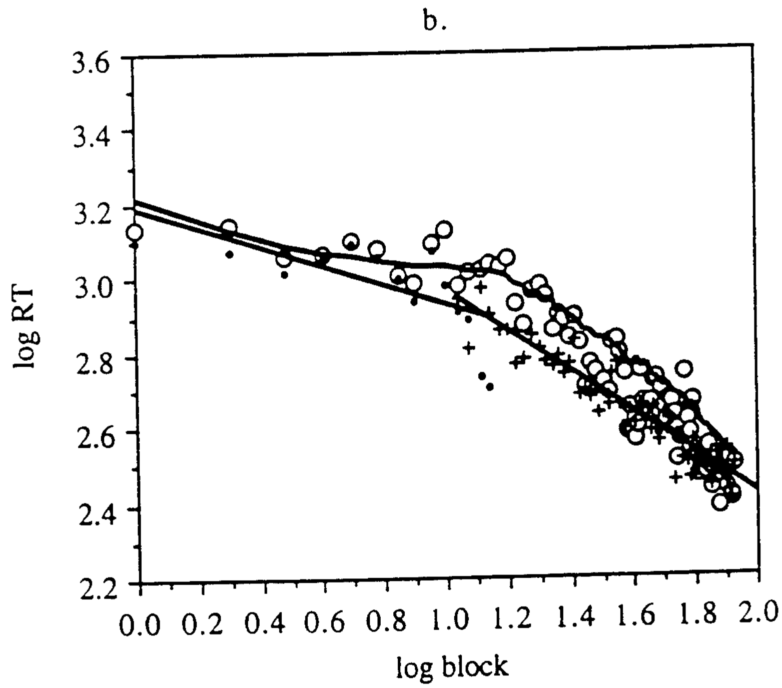


Figure 34. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 5 problems plotted as a function log block for both the overall data, and separately for the two strategies (algorithm or retrieval). Thin lines represent best fitting power function fits to each strategy, and thick lines represent CPL fits to the overall data (Experiment 2).

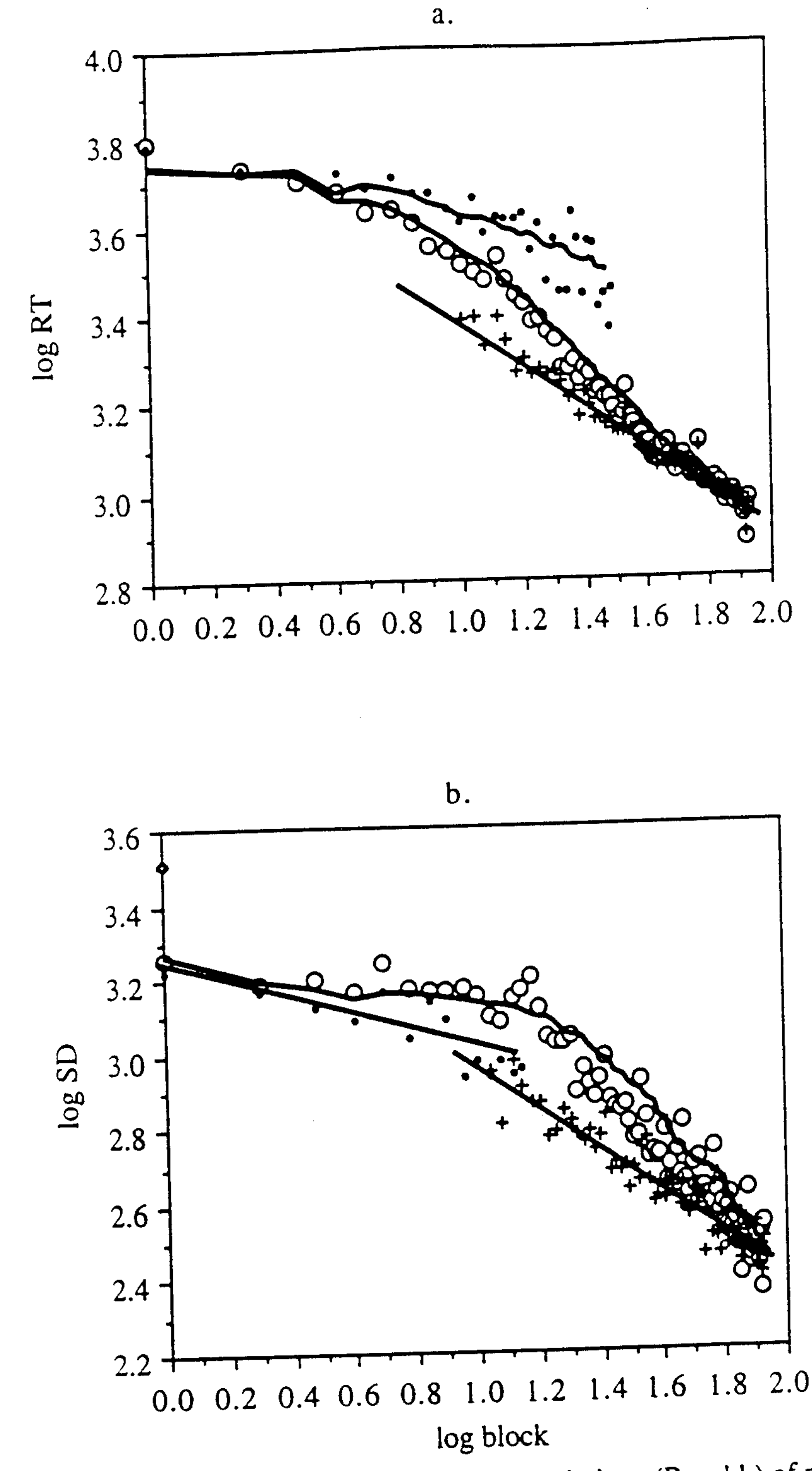


Figure 35. Log means (Panel a) and standard deviations (Panel b) of reactions times for addend 7 problems plotted as a function log block for both the overall data, and separately for the two strategies (algorithm or retrieval). Thin lines represent best fitting power function fits to each strategy, and thick lines represent CPL fits to the overall data (Experiment 2).

CHAPTER 4

GENERAL DISCUSSION

Two models of tasks exhibiting a transition from algorithm-based to retrieval-based performance were tested; a model derived from the instance theory of automaticity (Logan, 1988), and the CPL model. The RT and SD results of both experiments were clearly inconsistent with the current version of the instance theory; the power law did not hold overall for either the RT and SD data, and the rate parameters for the RT and SD were reliably different. In contrast, the CPL model fit almost exactly to the data from both experiments. This strong empirical support for the model invites further exploration of the possible implications of the model and of its theoretical underpinnings for a variety of related issues. In the following sections, I will consider potential implications for the topics of automaticity, mechanisms of learning in skill acquisition, representation in memory, and empirical laws of practice.

The CPL Model and Issues in Automaticity

The instance theory makes the point that much of what is termed "automatic" can be understood as a postattentional, rather than preattentional, phenomenon (Logan, 1992). Logan (1988) argues that retrieval from memory is a postattentional automatic process which is qualitatively distinct from other automatic phenomena which are considered to be preattentional (Shiffrin & Schneider, 1977; Triesman, Vieira, & Hayes, 1992). There is currently much discussion in the literature about whether either or both of these classes of phenomena should be labeled automaticity (Bargh, 1992; Logan, 1992; Treisman et al., 1992), but these discussions are beyond the scope of the current paper. For current purposes, I will only note that the CPL model, like the instance theory, is a memory-based theory of "automaticity". It thus inherits the advantages of the instance theory relative to purely process-based theories which assume that the processes underlying a task do not change qualitatively with practice, but just become more efficient, or automatized. However, the CPL model also predicts speedup in the algorithm, whereas the instance theory assumes no change in algorithm execution times. There is now ample evidence that algorithm speedup does occur, in the pound arithmetic task (see Experiment 1), in the alphabet arithmetic task (see Logan & Klapp, 1991), and in other related tasks (Carlson & Lundy, 1992).

Memory-based theories of automaticity such as the instance theory and the CPL model imply that the problem of diagnosing when a task is automatized amounts to determining when the transition to single-step retrieval has occurred. Two common approaches are to collect protocols, and to look for specificity of learning by conducting transfer tests (Experiments 1 of this paper; Klapp et al., 1991). If algorithms of differing difficulty levels are used, then slopes of mean RTs as a function of algorithm difficulty can also provide a useful diagnostic: when the slope approaches zero one can infer that automaticity has been achieved (Klapp et al., 1991; Shiffrin & Schneider, 1977). The CPL model suggests two additional techniques which should provide useful supplements to these standard techniques. First, it predicts that systematic and predictable deviations from linearity in log-log plots of the RT and STD data will occur during the development of automaticity. These deviations will be more extreme for the SD data, suggesting that this variable will provide a more sensitive index. Second, if strategy protocols are collected, the model predicts log-log linear speedup and reduction in SD within each strategy, and this prediction appears to be particularly robust for the retrieval strategy. The presence of separate power functions not only provides evidence of automaticity in its own right, but it also validates the protocol data, which can in turn be used as a rough estimate of the degree to which automaticity has been achieved at various points during practice.

There are at least two limitations which would need to be heeded, however, if these techniques are used to diagnosis automaticity. First, if the component power functions for the two strategies have the same parameter values, then no deviations from the power function would be expected in the overall data. This state of affairs is likely to be rare, but it should be investigated before inferring either that automaticity has not been achieved or that the CPL model does not hold. Second, any deviations from linearity in the overall data may be difficult to detect when the algorithm strategy is only marginally more time consuming than the retrieval strategy. The addend 2 data of Logan (1988) is an example in point. This limitation is not trivial as some domains of theoretical interest, such as the keyword method of foreign vocabulary learning (Crutcher, 1989), and stroop training effects (Clawson, 1994), are likely to exhibit very subtle transition effects. Experiments exploring such tasks would likely require very many observations or excellent control of noise, or both, for any existing deviations from the power law to be observable.

Implications for Mechanisms of Learning in Skill Acquisition

Several recent studies exploring complex mental arithmetic tasks (e.g., multi-column addition) have demonstrated that knowledge "restructuring" is an important consequence of practice (Carlson & Lundy, 1992; Charness & Campbell, 1988; Frensch & Geary, 1993). These restructuring effects are typically interpreted in terms of knowledge compilation (Anderson, 1983), which in turn reflects two component processes; proceduralization and composition. Proceduralization is a process whereby the need to retrieve declarative knowledge from long-term memory to execute a production is bypassed by creating a new production which applies directly to specific situations. Composition is a process whereby two or more productions are combined into one production which can more quickly and accurately perform the actions of the original two productions.

It does not appear that either proceduralization or composition corresponds very well to the primary type of learning which is occurring in this task domain.

Proceduralization may be occurring within the algorithm during practice, and it may account for much of the speedup on the algorithm, but it is not clear how it could account for the transition from algorithm to retrieval. Composition also does not seem appropriate for two reasons. First, it is generally assumed to operate on adjacent productions (Anderson, 1983), and thus multiple hierarchical compositions of a multi-step algorithm across several trials would be required before performance could mimic direct retrieval. This sort of progressive compilation does not seem to characterize these tasks. Indeed, this formulation of composition effects predicts power law speedup (Anderson, 1983), which was not observed. Second, a transition to retrieval-like performance via a composition mechanism would preserve much information in the composed production which is simply unnecessary in this case. For example, a composed production for pound arithmetic would require retrieving all of the arithmetic facts corresponding to the steps of the algorithm in one production, even though only the result of the last step, the final answer, would be needed for output. An associative learning mechanism allowing for direct retrieval (i.e., a literal bypassing of intermediate steps) is much simpler and would likely allow for faster performance. Note, however, that composition is a reasonable learning mechanism in other situations where each of the productions to be composed produces necessary output. For example, it is plausible that individual productions for dialing each digit of a frequently used telephone number are composed into a single production for dialing that number (see Anderson, 1983).

Implications for Representation in Memory

The good fits of the logistic strength-threshold model to the strategy probing data from both experiments make the point that strength representation is viable in theories of this task domain. In contrast, the current version of the instance theory cannot account for the results of these experiments. This is not to say that no instance-based approach could conceivably work. Rather, a satisfactory instance model remains to be demonstrated.

Arguments for strengthening are most compelling once the ability to retrieve consistently the answer from memory has been established for a given problem. But what about the effects of repetition before the transition to retrieval has occurred? The CPL model assumes that strengthening process occurs throughout practice regardless of which strategy is used to produce the answer. But there are other possibilities. It is possible that strengthening requires either recall of the answer, or conscious recognition of the problem-answer association after the answer is produced using the algorithm. Under this scenario, any existing memories would be instance-based until one of these instances is recalled or recognized on a subsequent trial. That instance could then be strengthening with additional practice. Note that the ability to recognize the problem-answer combination might occur well before the ability to recall the answer has been achieved (e.g., it might occur by the second block of practice), and thus the purely strength-based memory assumption of the CPL model might still provide a very good approximation to the underlying memory structure throughout most of the practice interval. Another factor which needs to be considered in modeling the effects of performance on the memory representation is the generation effect (Slamecka & Graf, 1978): in most circumstances, recall has a more pronounced positive effect on subsequent performance than does recognition. This effect suggests that there may be a marked discontinuity in memory strength for a given item after the first successful recall of the answer from memory. This possibility is consonant with the finding that a large portion of the transitions to retrieval at the item level approximate a step function; once the first recall occurs, strength may be incremented enough that it will nearly always be above threshold for that item.

These considerations highlight the fact that the simple strengthening assumptions of the CPL model are likely to be oversimplified. Nevertheless, the general notion of strengthening is not inconsistent with any of these considerations, and it appears to be a promising starting point for development of a more complete model of memory processes in this task domain.

Implications for the Power Law of Practice

Data from two experiments showed that the power law does not hold for tasks that exhibit a single-step strategy transition from algorithm-based to memory-based performance. Given the apparent ubiquity of the power law in other data sets, however, it is important to determine whether the failure to fit the data in these experiments represents a genuine failure of the law, or rather can be attributed to the particular form of the power function which was applied. The generalized form of the power law proposed by Newell and Rosenbloom (1981) includes four parameters; the intercept and slope parameters (used in the CPL fits to the component strategies), an asymptote parameter described in Chapter 1 and used in the instance theory fits, and a previous learning parameter which accounts for any previous experience related to the task. This generalized power function takes the form,

$$RT = a + b*(N - c)^{-d},$$

where a represents asymptotic RT, b represents RT on the first trial of the experiment, c represents the number of learning trials prior to the experiment, and d is the rate parameter.

There is no doubt that the simple two-parameter power function which ignores the asymptote and previous learning does not provide a good account of the data from Experiments 1 and 2. Adding the asymptote, a, also does not yield good fits, as indicated by the instance theory fits which included this parameter. However, addition of a previous learning parameter, c, does substantially improve the fits. Indeed, much of the deviation from the power function in the RT data can be overcome by fitting a separate prior learning parameter to each of the RT curves from Experiments 1 and 2. Thus, one might argue that the data are consistent with the more general version of the power law. This interpretation, however, runs into several serious difficulties. First, the pound arithmetic and alphabet arithmetic tasks are both tasks with which subjects surely had no experience at the outset of the experiment. Thus, if previous experience is defined in terms of experience with the task as defined in the experiment, then the previous learning parameter must take a value of zero. A more liberal interpretation of the previous learning parameter is that it reflects any previous learning which is relevant to performing the experimental task, such that nonzero values can occur even when the task proper is novel. This approach is reasonable in principle, and in effect predicts that previous learning should take on positive values for virtually <u>any</u> task environment (because some general previous learning will always be relevant). However, there are both conceptual and empirical difficulties with this assumption. First, defining previous learning in this generic way effectively removes any constraints on what values of the parameter c are reasonable in a given context; by what criteria could one evaluate whether an obtained value for c is reasonable if previous learning is not tied in some explicit way to the experimental task? Without such a constraint, the four-parameter power law is extremely powerful, and may even prove unfalsifiable given the noise which is inherent in any data set.

A myriad of empirical inconsistencies brought about by this liberal interpretation of previous learning is even more problematic. Consider first the fact that 9 of the 18 data sets fit with the four parameter power function by Newell and Rosenbloom (1981) yielded zero as the best fitting value of the parameter c. These parameter estimates would have been negative had c not been bounded in the fits to have nonnegative values. Further, there is no evidence in their analyses that the positive values of c

obtained for the other 9 data sets yielded any visually or statistically meaningful improvements in the fits. If a generic type of previous learning was not implicated in these other data sets, why should it play such an important role in the current experiments?

Even more compelling evidence that previous learning effects do not account for the deviations from the power functions in the current experiments can be garnered by considering some arithmetic data sets which were collected more recently. Rickard (1992) explored adult performance on simple multiplication (4 x 7) and division (32 ÷ 8) problems. It is difficult to conceive of a task which would intuitively be more likely to exhibit previous learning effects. Yet, for both arithmetic operations, the previous learning parameter took values of zero. Next consider a complex arithmetic task studied by Carlson and Lundy (1992). Although they did not fit generalized power functions to their data, the two-parameter power function provided visually good fits in a varied data condition which precluded any strategy transitions with practice (see Figure 4). It did not provide a good fit, however, to the consistent data condition, in which the CPL model predicts that strategy transitions will occur. Finally, consider the nontransition subject of the present Experiment 1. This individual's performance was initially faster than all but one of the 18 subjects who did show a transition. Thus, relevant previous learning for this individual would presumably be at least as great as that of the other 18 subjects on average. Nevertheless there was no deviation from log-log linearity in this subject's data. In sum, these findings raise a serious question about a possible previous learning account of the deviations from the two-parameter power function evident in the current experiments: Why would previous learning show itself so clearly when strategy transitions occur but not be evident whatsoever in a variety of other strongly analogous learning contexts where such transitions do not occur?

An additional problem with a previous learning interpretation is related to the effect of addend size in Experiment 2. Any general previous learning which might be operating in the alphabet arithmetic task should be independent of something as specific as the addend size of the algorithm. Yet, the fact that the deviations from the power function increase with increasing addend size dictates that different values of *c* would be needed to provide reasonable power function fits to these data. Comparison of Logan's (1988) results for addend size 2 problems with the current results for addend size of 5 and 7 underscores this difficulty. For addend 2 problems, there is almost no deviation from log-log linearity, whereas for addend sizes of 5 and 7, the deviations are clear. A similar difficulty would plague any attempt to generalize the power function to the SD data from either experiment. Any previous learning must be exactly the same when fitting the mean RT and the SD of the RT of a given data set. However, the functional characteristics of the deviation from the power function were very different for the RT and SD data, such that a single previous learning parameter would not be able to correct both deviations simultaneously. For example, it was found that the optimal value of the previous learning parameter for addend 7 RTs in Experiment 2 to is 9.09. The value for SDs is 76.20.

All of these arguments in combination make a previous learning account of the deviations from the power law in the current experiments untenable. In contrast, the strategy specific power function speedup assumed in the CPL model accounts naturally for these results, and also does not introduce empirical inconsistencies when other data sets are considered. According to the model, other data sets which do not exhibit deviations from the power function reflect either increases in efficiency of execution of a given strategy, or different types of transitions such as proceduralization or composition which might still yield power function speedup (see Anderson, 1983).

Reinterpreting the Power Law

Some elaborated version of the power law appears to be needed. I will make two speculative proposals. First, I propose that the previous learning should by default be assumed to be zero unless the exact task as defined in the laboratory is known to have been performed previously (this is typically done implicitly as researchers rarely include previous learning parameters in power function fits). This proposal is motivated purely by the lack of any empirical evidence which demonstrates the need for nonzero previous learning otherwise. It is unclear at present why general previous learning is not a factor empirically. Because the power function is translation dependent, it strictly must be the case that any relevant previous learning will have some degree of impact. One possibility as that improvements in performance on virtually any task is dominated by the aspects of the task which are truly novel (e.g., learning to execute a complex cognitive algorithm). This hypothesis might even explain the estimate of zero previous learning for adults on simple arithmetic tasks (Rickard, 1992). Rickard and Bourne (1994) showed that most speedup with practice in their experiments was specific to perceptual and motor characteristics of the task which were novel to the subjects, although there was also some nonspecific speedup attributable to "cognitive stage" arithmetic fact retrieval processes.

A second proposal addresses the issue of how to define the power law so that it will hold in any task environment. The obvious working hypothesis is that the power law always holds within strategies, but will not necessarily hold when there are strategy shifts. One problem which this hypothesis raises is determination of what constitutes a unique strategy. A workable solution is to define a strategy in terms of the reportable contents of working memory, which in turn reflects the cognitive steps involved in processing (Ericsson & Simon, 1993). Any change in these cognitive steps as indicated by protocols constitutes a strategy shift. Whenever these changes are not observed, the power law should hold. When they are observed, the law would not be guaranteed to hold.

The difficulty with the definition proposed above is that there are clearly cases in which qualitative shifts occur which would constitute strategy shifts by a protocol criterion, but for which the power law nevertheless does hold. For example, proceduralization effects appear to preserve power function speedup (e.g., Neves & Anderson, 1981). This fact suggest the alternative proposal that the law holds unless specific types of strategy shifts occur. Processes such as chunking, proceduralization, and composition might generally produce power law speedup. Other processes, like the associative learning processes discussed in this paper, do not always produce the power law. This possibility suggests the hypothesis that non-strategic or "automatic" learning of any type will produce power law speedup, but that strategic efforts to adopt a new strategy in some cases will not. There is some evidence that strategic attempts to acquire the problem-answer associations are important to making the transition to retrieval (e.g., Logan, 1988). More research is needed to test this possibility.

The results of both experiments also support the CPL assumption that reduction in SD follows a power function within a given strategy. Clear power function reduction in SD was also observed by Rickard and Bourne (1994) on a simple arithmetic fact retrieval task (see Figure 1) and by Logan (1988) on a lexical decision task which also probably does not involve strategy transitions. The power function clearly does not, however, provide a good account of reduction in SD with practice when clear strategy transitions take place.

One might speculate that the current demonstration that the power law does not always hold is a relatively minor "exception to the rule" which may not replicate beyond the current task domain. This hypothesis may be correct, but it should be considered with caution. This exception to the rule may prove to be fairly common now that a general theoretical account has been formulated. The CPL model suggests that deviations from power functions which are typically overlooked (Newell & Rosenbloom, 1981) or judged inconclusive (Carlson & Lundy, 1992; Logan, 1988) might be more profitably interpreted as reflecting strategy shifts. It is also easy to see the potential theoretical consequences of this demonstration by considering that the instance theory (Logan, 1988) was developed largely in an effort to account for the power law, which was simply understood at that time to be true in any situation. The current results suggest that theories which do not predict the power law should not be dismissed out of hand, especially when those theories address task domains which exhibit marked strategy shifts with practice.

REFERENCES

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). Rules of the mind. Hillsdale, NJ: Lawrence Erlbaum.

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. Cognition, 44, 75 - 106.

Bargh, J. A. (1992). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. American Journal of Psychology, 105, 181-189.

Battig, W. F. (1979). The flexibility of human memory. In L. S. Cermak & F. I. M. Craik (Eds.), Levels of processing and human memory (pp. 23-44). Hillsdale, NJ: Erlbaum.

Campbell, J. I. D., & Oliphant, M. (1992). Representation and retrieval of arithmetic facts: A network-interference model and simulation. In Campbell, J. I. D. (editor), The nature and origins of mathematical skills (pp. 331- 364). North Holland: Elsevier.

Carlson, R. A., & Lundy, D. H. (1992). Consistency and restructuring in learning cognitive procedural sequences. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 127-141.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15, 517-526.

Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. Journal of Experimental Psychology: General, 117, 115-129.

Clawson, D. M. (1994). Dimensions of practice specificity in the classic stroop color-word task. Unpublished Doctoral Dissertation, University of Colorado, Boulder, Colorado.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. Memory & Cognition, 19, 151 - 158.

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. Ergonomics, 2, 153-166.

Crutcher, R. J. (1989). The Role of Mediation in Knowledge Acquisition and Retention: Learning Foreign Vocabulary Using the Keyword Method. Unpublished doctoral Dissertation. University of Colorado, Boulder, Colorado

Cohen, J. D. , Dunbar, K. & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the stroop effect. Psychological Review, 7, 332 - 361.

Ericsson, K. A. (1985). Memory Skill. Canadian Journal of Psychology, 39, 188-231.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press

Fendrich, D. W., Healy A. F., & Bourne, L. E. Jr. (1993). Mental arithmetic: Training and retention of multiplication skill. In C. Izawa (Ed.), Cognitive psychology applied (pp. 111-133). Hillsdale, NJ: Erlbaum.

Frensch, P. A., & Geary, D. C. (1993). Effects of practice on component processes in complex mental addition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 1-24.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), The psychology of learning and motivation (pp. 47 - 91). New York: Academic Press.

Klapp, S. T., Boches, C. A., Trabert, M. A., & Logan, G. D. (1991). Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved? Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 196 - 209.

Kolers, P. A. (1975). Memorial consequence of automatized encoding. Journal of Experimental Psychology: Human Learning and Cognition, 1, 689-701.

Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492 - 527.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 883 - 914.

Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? Journal of Experimental Psychology: Learning, Memory, and Cognition, 17, 179 - 295.

MacKay D. G. (1982). The problem of flexibility, fluency, and speed-accuracy trade-off in skilled behavior. Psychological Review, 89, 483-505.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of the effect of context in perception, Part 1. An account of basic findings. Psychological Review, 88, 375-407.

McCloskey, M., Caramazza, A. & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. Brain and Cognition, 4, 171-96.

Moran, T. P. (1980). Compiling cognitive skill. AIP Memo 150, Xerox PARC.

Neves , D. M. & Anderson, J. R. (1981). Knowledge compilation: Mechanisms for the automatization of cognitive skills. In J. R. Anderson (Ed.), Cognitive skills and their acquisition, Hillsdale, N.J.: Lawrence Erlbaum Associates.

Newell, A. (1993). Unified theories of cognition. Harvard University Press: Cambridge, MA.

Newell, A, & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (Ed), Cognitive skills and their acquisition. Hillsdale, New Jersey: Lawrence Erlbaum.

Pirolli P. L & Anderson, J. R. (1985). The role of practice in fact retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 136-153.

Rickard, T. C. (1992). Acquisition and transfer of mental multiplication skill. Unpublished masters thesis, University of Colorado, Boulder, Colorado.

Rickard, T. C., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Some tests of an identical elements model. Unpublished manuscript.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (in press). On the cognitive structure of basic arithmetic skills: Operation, order, & symbol transfer effects. Journal of Experimental Psychology: Learning, Memory, & Cognition.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. Behavior Research Methods, Instruments, and Computers, 20, 206-217.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson, R. J. Hernstein, G. Lindzey, & R. D. Luce (Eds.) Steven's handbook of experimental psychology: Vol. 2 Learning and Cognition (2nd ed. pp. 739-811). New York: Wiley.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending and a general theory. Psychological Review, 84, 127-190.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill.Journal of Experimental Psychology: General, 117, 258-275.

Siegler, R. S. & Jenkins, E. (1989). How children discover strategies. Lawrence Erlbaum: Hillsdale, N.J.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. Journal of Experimental Psychology: Human Learning and Memory, 4, 592-604.

Townsend, J. T., & Ashby, F. G. (1984). Stochastic modeling of elementary psychological processes. Cambridge, MA: Cambridge University Press.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. American Journal of Psychology, 105, 341-362.

Zbordoff N. J. & Logan, G. D. (1986). On the Autonomy of Mental Processes: A Case Study of Arithmetic. Journal of Experimental Psychology: General, 115, 118-130

APPENDIX A

PROBLEM SETS USED IN EXPERIMENT 1

Set 1

3 # 17 = __
4 # 12 = __
5 # 16 = __
6 # 19 = __
7 # 15 = __
8 # 13 = __
3 # __ = 20
4 # __ = 29
5 # __ = 34
6 # __ = 18
7 # __ = 12
8 # __ = 27

Set 2

3 # __ = 32
4 # __ = 21
5 # __ = 28
6 # __ = 33
7 # __ = 24
8 # __ = 19
3 # 11 = __
4 # 16 = __
5 # 19 = __
6 # 18 = __
7 # 12 = __
8 # 17 = __

Set 3

3 # 17 = __
4 # 12 = __
5 # 16 = __
6 # 19 = __
7 # 15 = __
8 # 13 = __
3 # __ = 34
4 # __ = 11
5 # __ = 30
6 # __ = 25
7 # __ = 32
8 # __ = 21

Set 4

3 # __ = 32
4 # __ = 21
5 # __ = 28
6 # __ = 33
7 # __ = 24
8 # __ = 19
3 # 18 = __
4 # 11 = __
5 # 17 = __
6 # 15 = __
7 # 19 = __
8 # 14 = __

Set 5

3 # 11 = __
4 # 16 = __
5 # 19 = __
6 # 18 = __
7 #12 = __
8 # 17 = __
3 # __ =
4 # __ =
5 # __ =
6 # __ =
7 # __ =
8 # __ =

Set 6

3 # __ = 20
4 # __ = 29
5 # __ = 34
6 # __ = 31
7 # __ = 18
8 # __ = 27
3 # 18 = __
4 # 11 = __
5 # 17 = __
6 # 15 = __
7 # 19 = __
8 # 14 = __

APPENDIX B

PROBLEMS USED IN EXPERIMENT 2

<u>True</u>	<u>False</u>
E + 3 = H	E + 3 = I
N + 3 = Q	N + 3 = R
H + 3 = K	H + 3 = L
K + 3 = N	K + 3 = O
J + 5 = O	J + 5 = P
G + 5 = L	G + 5 = M
P + 5 = U	P + 5 = V
M + 5 = R	M + 5 = S
L + 7 = S	L + 7 = T
I + 7 = P	I + 7 = Q
F + 7 = M	F + 7 = N
O + 7 = V	O + 7 = W